Toward Resolution-Invariant Person Reidentification via Projective Dictionary Learning

Kai Li, *Student Member, IEEE*, Zhengming Ding, *Member, IEEE*, Sheng Li, *Member, IEEE*, and Yun Fu, *Senior Member, IEEE*

***Abstract*— Person reidentification (ReID) has recently been widely investigated for its vital role in surveillance and forensics applications. This paper addresses the low-resolution (LR) person ReID problem, which is of great practical meaning because pedestrians are often captured in LRs by surveillance cameras. Existing methods cope with this problem via some complicated and time-consuming strategies, making them less favorable, in practice, and meanwhile, their performances are far from satisfactory. Instead, we solve this problem by developing a discriminative semicoupled projective dictionary learning (DSPDL) model, which adopts the efficient projective dictionary learning strategy, and jointly learns a pair of dictionaries and a mapping function to model the correspondence of the cross-view data. A parameterless cross-view graph regularizer incorporating both positive and negative pair information is designed to enhance the discriminability of the dictionaries. Another weakness of existing approaches to this problem is that they are only applicable for the scenario where the cross-camera image sets have a globally uniform resolution gap. This fact undermines their practicality because the resolution gaps between cross-camera images often vary person by person in practice. To overcome this hurdle, we extend the proposed DSPDL model to the variational resolution gap scenario, basically by learning multiple pairs of dictionaries and multiple mapping functions. A novel technique is proposed to rerank and fuse the results obtained from all dictionary pairs. Experiments on five public data sets show the proposed method achieves superior performances to the state-of-the-art ones.**

***Index Terms*— Dictionary learning, fusion, low resolution (LR), person reidentification (ReID), reranking.**

I. Introduction

CROSS-CAMERA pedestrian matching, formally referred as person reidentification (ReID), has been attached with increasing attention due to its importance in surveillance and forensics applications. For human identification, biometrics such as face and gait are widely exploited, due to their strong discrimination. However, these patterns are unreliable for ReID because of the arbitrary of human poses and low quality of images captured by surveillance cameras. Therefore, ReID mainly relies on the visual appearance of the human body by assuming that there are not fundamental appearance changes (say, wearing different clothes) of the same person in different camera views. Even underlaid on this assumption, ReID remains a very challenging problem, due to the variances in pose, resolution, illumination, occlusion, and so on. Existing methods solve this problem either by exploring invariant and discriminative features [1]–[6] or developing robust distance metrics [7]–[11].

This paper targets at addressing one of the key factors that impact ReID, i.e., large cross-view resolution gaps. Investigating this problem is of great practical meaning because a person is often captured in low resolutions (LRs) by real-world surveillance cameras, such that it is often required to perform LR-ReID [12]. Although impressive progress has been made for ReID in recent years and many effective methods have been proposed, these methods have only been proved to be effective when the cross-view images are of similar resolutions. For the case where the images are of great resolution divergences, the good performance cannot be guaranteed. There are several initial attempts targeting at the great resolution gap problem. The approach in [12] learns a shared subspace across different scales and a discriminative distance metric which minimizes a novel heterogeneous class mean discrepancy criterion. Wang *et al.* [13] proposed to learn a discriminating surface that separates scale-distance functions between images of the same persons and those of different persons, and use it for reidentifying persons. Wang *et al.* [14] proposed to extract an effective feature from low- and high-resolution (HR) pedestrian images by building two coupled marginalized denoising autoencoders. Jing *et al.* [15] proposed a semi-coupled low-rank discriminant dictionary learning (SLD^{2}L) method by dividing images into patches and learning semi-coupled dictionaries for corresponding image patch clusters. Despite of these sophisticated techniques, the performances of these methods are far from satisfactory, and some of them are too time consuming. Another common limitation of these methods is that they assume there is a globally uniform resolution gap between cross-view pedestrian images, i.e., the resolution gaps of the cross-view images of different persons are the same. This assumption undermines the

Manuscript received January 16, 2018; revised May 13, 2018; accepted September 24, 2018. Date of publication November 2, 2018; date of current version May 23, 2019. This work was supported in part by the NSF IIS Award under Grant 1651902 and in part by the U.S. Army Research Office Award under Grant W911NF-17-1-0367. *(Corresponding author: Kai Li.)*

K. Li is with the Department of Electrical and Computer Engineering, College of Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: kaili@ece.neu.edu).

Z. Ding is with the Department of Computer, Information and Technology, Indiana University—Purdue University Indianapolis, Indianapolis, IN 46202 USA (e-mail: zd2@iu.edu).

S. Li is with the Department of Computer Science, University of Georgia, Athens, GA 30602 USA (e-mail: sheng.li@uga.edu).

Y. Fu is with the Department of Electrical and Computer Engineering, College of Engineering, Northeastern University, Boston, MA 02115 USA, and also with the College of Computer and Information Science, Northeastern University, Boston, MA 02115 USA (e-mail: yunfu@ece.neu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2018.2875429

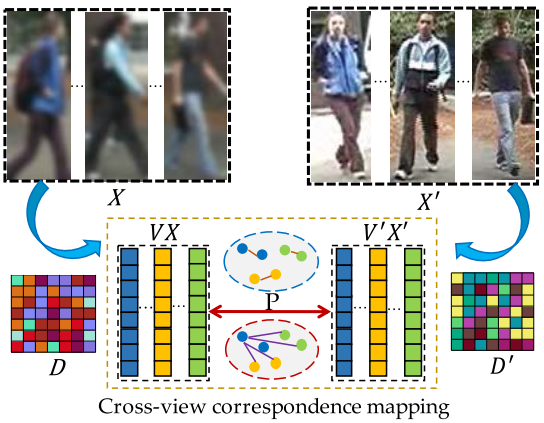


Fig. 1. Framework of the proposed model for LR-ReID. By utilizing labeled low- and HR pedestrian images X and X', we jointly learn a dictionary pair D and D', and a mapping function P, which relates the new codings of X and X'. V and V' are the two projections which map X and X' to their new codings VX and $V'X'$, respectively. Positive and negative pair information is incorporated in a cross-view graph regularization term to guide learning dictionaries of strong discriminability.

practicality of these methods because the resolution gaps of cross-camera images usually vary person by person in practice.

To avoid limitations of the existing methods, we propose a discriminative semicoupled projective dictionary learning (DSPDL) model to effectively and efficiently solve the LR-ReID problem. DSPDL adopts the efficient projective dictionary technique, and jointly learns a pair of dictionaries and a mapping function to model the correspondence of cross-view data. To enhance the discriminative power of the learned dictionaries, a novel parameterless cross-view graph regularizer is proposed to incorporate both positive and negative cross-view pair information. Fig. 1 shows the framework. We apply DSPDL on the LR-ReID problem, for both the uniform and variational resolution gap scenarios, by treating cross-camera pedestrian images as the cross-view data. For the variational resolution gap scenario, we propose to use DSPDL to learn multiple pairs of dictionaries and multiple mapping functions, and further formulate a novel technique to rerank and fuse the results obtained from all dictionary pairs. Our major contributions are summarized as follows.

- We propose to model the correspondence of cross-view data via SPDL. We devise a model to jointly learn cross-view dictionaries and a mapping function which establishes the correspondence of cross-view data. Through the introduction of the mapping function, the stringent correspondence between the cross-view data is relaxed, making it possible to maximize the feature representation ability of the learned dictionaries.
- We devise a novel cross-view graph regularizer which unifies positive and negative pair information in a parameterless fashion. The incorporation of discriminative information in the regularizer boosts the discriminability of the learned dictionaries, thus facilitating our model to distinguish correct person pairs from incorrect ones. The graph regularizer is parameter free, so that our method is supposed to have robust performances with images of great diversity.
- We extend the proposed DSPDL model to the variational resolution gap scenario by learning multiple pairs of dictionaries and multiple mapping functions. A novel technique is proposed to rerank and fuse the results obtained from all dictionary pairs.
- We evaluate our method on five benchmark data sets by comparing with the state-of-the-art approaches. The results show that our method achieves remarkable improvements.

This paper is a journal extension of our conference publication [16]. We make improvements by extending the proposed DSDPL model to the variational resolution gap scenario, and propose a novel fusion and reranking technique to achieve this. We also provide more mathematic analysis and experimental results to evaluate the proposed model for both uniform and variational resolution gap scenarios.

The rest of this paper is organized as follows. In Section II, we review the related works. The proposed DSPDL model and its application on LR-ReID is elaborated in Sections III and IV, respectively. The experimental results and analysis are presented in Section V. Section VI gives the conclusion.

II. Related Works

Person ReID can generally be classified into three categories: pedestrian description or feature learning-based methods, distance metric learning-based methods, and deep learning-based methods.

Based on the fact that a person in different camera views should be similar in appearance, many pedestrian image description-based methods have been proposed for ReID [17]–[22]. There are various types of image features, including the color features from different channels, i.e., RGB, CbCr, LAB, and so on; and texture feature extracted by local binary patterns, histogram of oriented gradients, and scale-invariant feature transform descriptors. Since a single image descriptor is often not powerful enough to encode all the information that are essential for pedestrian image matching, concatenating the feature vectors of several image descriptor is commonly used. Apart from developing sophisticated appearance description techniques, another line of methods in this category focus on learning discriminative cross-view features based on dictionary learning [15], [23]–[25]. The basic idea is to learn the dictionary pair under which the cross-view images of the same person have similar feature representations. Besides using low-level color and texture features, another good choice is the attribute-based features which can be viewed as mid-level representations. It is believed that attributes are more robust to image transformations compared to low-level features [26]–[29].

The second category is the distance metric learning-based methods. The general idea of metric learning-based ReID methods is to learn some distance metrics under which the vectors of the same identities are pushed closer while the vectors of different identities are pulled further apart. Keep it simple and straightforward metric (KISSME) [30] is one of the most acknowledged methods in this category, which decides whether a pair of description vectors

are similar or not by formulating it as a likelihood ratio test. The pairwise difference is employed and the difference space is assumed to be a Gaussian distribution with a zero mean. Inspired by KISSME, many metric learning-based ReID algorithms have been proposed, including regularized PCCA [7], local Fisher discriminant analysis (LFDA) [9], Information-Theoretic Metric Learning + [31], cross-view quadratic discriminant analysis (XQDA) [10], metric learning with accelerated proximal gradient (MLAPG) [32], and so on.

The third category is deep learning-based methods. Although deep learning has shown extraordinary advantages in many visual learning tasks, the lack of training data becomes the major bottleneck of applying deep learning for ReID. Most ReID data sets provide only two images for each identity such that they are insufficient to train complex deep learning models. For this reason, deep learning-based ReID methods are often unable to outperform the traditional methods [21], [33], [34]. It is also for this reason that many deep learning-based ReID methods focus on the Siamese model, in which two or more identical subnetworks share parameters during the training stage [35]–[38], thus reducing the parameters to be learned and accordingly are less demanding for the labeled data. The key drawback of Siamese model-based ReID methods is that they exploit only the pairwise identity labels (whether an image pair belongs to the same identity) of the ReID annotations, which does not make full use of the ReID labels. To break this limitation, some methods solve the ReID problem from the perspective of classification [39], [40] and pretrain their models on large classification data sets, say Imagenet [41], and finetune them on ReID data sets.

Our method solves the ReID problem also from the perspective of feature learning. Unlike existing dictionary learning-based ReID methods, we adopt the powerful projective dictionary learning technique, which avoids to solve l_1-norm optimization, thus making our method highly efficient. Meanwhile, we approach the LR problem by learning a mapping between low- and HR images along with the dictionaries, and define a novel parameterless cross-view graph regularizer to incorporate both positive and negative pair knowledge to enhance the discriminability of the learned dictionaries.

III. DSPDL Model

Denoted by $X \in \mathbb{R}^{d\times n}$ and $X' \in \mathbb{R}^{d\times n}$ two cross-view data sets, in which there exists one-to-one cross-view correspondence, i.e., $x_i \in X$ and $x_i' \in X'$ are the representations of the ith instance from the two views. For LR-ReID, X and X' correspond to the low- and high-pedestrian image sets, respectively; x_i and x_i' are the low- and HR images of the ith person. A cross-view projective dictionary learning (CPDL) framework can be formulated as

$$\begin{aligned}\min_{D,D',V,V'} \quad & \|X - DVX\|_F^2 + \|X' - D'V'X'\|_F^2 \\ & + \lambda_1 \Omega(V, X, V', X') \\ \text{s.t. } & \|d_i\| \le 1,\ \|d_i'\| \le 1, \quad i = 1, \ldots, k \end{aligned} \tag{1}$$

where $\|X - DVX\|_F^2$ and $\|X' - D'V'X'\|_F^2$ are the data fidelity terms which measure how well the cross-view data are expressed by the dictionaries, while $\Omega(V, X, V', X')$ ensures similar new codings of cross-view data in correspondence.

For the data fidelity terms, we adopt the recently proposed projective dictionary learning technique, which obtains the new codings of input features through analytical feature projection, i.e., the new coding of X (X') under D (D') is VX ($V'X'$), projected from X (X') by projection V (V') $\in \mathbb{R}^{k\times d}$. In this way, we avoid to solve an inefficient l_1-norm optimization problem [42], [43] because we do not need to add sparse constraint on the new codings of input features.

Projective dictionary learning was originally designed for classification [44], [45]. Here, we introduce it for person ReID. The main difference is that when it is applied for classification, dictionaries are learned for each class. We instead learn only one dictionary for all images from one camera view. For classification, the goal is to learn a dictionary which encodes discriminative information of each class. For ReID, on the other hand, since each person may only have several training samples across different views, it is not enough to learn a dictionary per person. It is more like a verification problem. So the goal is to learn dictionaries to ensure images of the same identity to have similar codings across different views. Therefore, the interperson and intraperson relationship is more emphasized in our model.

$\Omega(V, X, V', X')$ regularizes to learn dictionaries under which cross-view correspondences have similar new codings. Intuitively, we can push close the new codings of cross-view correspondences and set the regularizer as $\|VX - V'X'\|_F^2$. However, in some cases, the divergence between cross-view correspondences x_i and x_i' are too large such that the generalization power of the learned dictionaries shall be diminished if we directly push close the new codings. For example, for LR-ReID, the same person could vary a lot in resolutions under different camera views, leading to significant appearance disparities in the images. Learning dictionaries to overfit the training samples shall bring about generalization problems when applying the learned dictionaries on the test data set. To avoid this, we propose to jointly learn a mapping function P, which aims to bridge the large cross-view divergences, along with the dictionaries. We formulate an SPDL model as

$$\begin{aligned}\min_{D,D',V,V',P} \quad & \|X - DVX\|_F^2 + \|X' - D'V'X'\|_F^2 \\ & + \lambda_1 \|VX - PV'X'\|_F^2 + \lambda_2 \|P\|_F^2 \\ \text{s.t. } & \|d_i\| \le 1,\ \|d_i'\| \le 1, \quad i = 1, \ldots, k. \end{aligned} \tag{2}$$

The mapping function introduces flexibility for the cross-view correspondence, thus preventing the over-fitting problem.

One may have observed that SPDL only exploits the positive pair information to constrain on the learned dictionaries to output similar codings for cross-view correspondences. However, it neglects the negative pair information that shall be helpful to enhance the discriminative power of the learned dictionaries, i.e., outputting different codings for noncorresponding data samples. To remedy this, we construct two graphs, i.e., intrainstance cross-view graph G_s and interinstance cross-view graph G_d, which can be encoded by affinity matrices W_s and W_d, respectively, for samples from the views. Unlike the graphs constructed for classification tasks [46], [47], we focus on cross-view discriminative dictionary learning and consider only the edges between cross-view nodes while neglect the

intraview ones

$$W_s = \begin{bmatrix} 0 & W_a \\ W_a^\top & 0 \end{bmatrix}, \quad W_d = \begin{bmatrix} 0 & W_b \\ W_b^\top & 0 \end{bmatrix} \tag{3}$$

where

$$W_a^{ij} = \begin{cases} 1, & \text{if } y_i = y'_j \\ 0, & \text{otherwise;} \end{cases} \quad W_b^{ij} = \begin{cases} \dfrac{1}{n}, & \text{if } y_i \neq y'_j \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $y_i \in Y$ is the label of the ith sample from X, $y'_j \in Y'$ is the label of the jth sample from X', and n is the number of training samples. Let $Z = [VX, PV'X'] = \{z_1, z_2, \ldots, z_n, z_{n+1}, \ldots, z_{2n}\}$ be the collection of the new codings of X and X'. Our goal is to maximize the cross-view intrainstance similarity, while minimizing the cross-view interinstance similarity. Thus, we have the following formulations:

$$\begin{aligned} \max \sum_{i,j} z_i W_s^{ij} z_j^\top &= \operatorname{tr}(ZW_sZ^\top) \\ &= \operatorname{tr}\left([VX, PV'X'] \begin{bmatrix} 0 & W_a \\ W_a^\top & 0 \end{bmatrix} [VX, PV'X']^\top\right) \\ &= \operatorname{tr}((PV'X')W_a^\top (VX)^\top + (VX)W_a(PV'X')^\top) \end{aligned} \tag{5}$$

and

$$\begin{aligned} \min \sum_{i,j} z_i W_d^{ij} z_j^\top &= \operatorname{tr}(ZW_dZ^\top) \\ &= \operatorname{tr}\left([VX, PV'X'] \begin{bmatrix} 0 & W_b \\ W_b^\top & 0 \end{bmatrix} [VX, PV'X']^\top\right) \\ &= \operatorname{tr}((PV'X')W_b^\top (VX)^\top + (VX)W_b(PV'X')^\top) \end{aligned} \tag{6}$$

where $\operatorname{tr}(\cdot)$ is the trace operation of a matrix. Combining (5) and (6), we obtain

$$\begin{aligned} &\Omega(V, X, V', X', P) \\ &\quad = \operatorname{tr}((PV'X')W_b^\top (VX)^\top + (VX)W_b(PV'X')^\top) \\ &\qquad - \operatorname{tr}((PV'X')W_a^\top (VX)^\top + (VX)W_a(PV'X')^\top). \end{aligned} \tag{7}$$

This crafted graph regularizer unifies cross-view intrainstance similarity and cross-view interinstance dissimilarity constraints, and meanwhile considers the great disparities across different views. No parameter is introduced, and thus, it is expected that the proposed model will have robust performances on diverse scenarios. With the graph regularizer, we reach our DSPDL model

$$\begin{aligned} \min_{D,D',V,V',P} \; & \|X - DVX\|_F^2 + \|X' - D'V'X'\|_F^2 \\ & + \lambda_1 \Omega(V, X, V', X', P) + \lambda_2 \|P\|_F^2 \\ \text{s.t. } & \|d_i\| \leq 1, \; \|d'_i\| \leq 1, \quad i = 1, \ldots, k \end{aligned} \tag{8}$$

where $\Omega(V, X, V', X', P)$ is defined in (7).

A. Optimization

To facilitate the optimization of our proposed DSPDL model in (8), we introduce three relaxation variables A, B', and B, and use them to replace VX, $V'X'$, and $PV'X'$, respectively. In this way, we rewrite (8) as

$$\begin{aligned} \min_{\substack{D,D',V,V' \\ A,B,B',P}} \; & \Phi(D, D', V, V', A, B, B', P) \\ & = \|X - DA\|_F^2 + \|X' - D'B'\|_F^2 \\ & \quad + \lambda_1 \Omega(A, B) + \lambda_2 \|P\|_F^2 + \alpha \|A - VX\|_F^2 \\ & \quad + \alpha \|B' - V'X'\|_F^2 + \beta \|B - PB'\|_F^2 \\ \text{s.t. } & \|d_i\| \leq 1, \; \|d'_i\| \leq 1, \quad i = 1, \ldots, k \end{aligned} \tag{9}$$

where $\Omega(A, B) = \operatorname{tr}(BW_b^\top A^\top + AW_bB^\top) - \operatorname{tr}(BW_a^\top A^\top + AW_aB^\top)$, and $\alpha = \beta = 10^{-3}$ are small penalty parameters.

The variables in (9) can be optimized one by one via fixing the others when optimizing one [44], [48]. The step-by-step optimization procedures are as follows.

1) Update A: By keeping only the terms relevant to A, we obtain $\min_A \Phi(A) = \|X - DA\|_F^2 + \lambda_1 \Omega(A, B) + \alpha \|A - VX\|_F^2$. Let the derivative of Φ with respect to A be zero, i.e., $(\partial \Phi / \partial A) = 0$, we have the closed-form solution of A as

$$A = (D^\top D + \alpha I)^{-1}(D^\top X + \alpha VX + \lambda_1 BW_a^\top - \lambda_1 BW_b^\top). \tag{10}$$

2) Update B: Ignoring irrelevant terms with respective to B, the objective function reduces to $\min_B \Phi(B) = \lambda_1 \Omega(A, B) + \beta \|B - PB'\|_F^2$. Setting $(\partial \Phi / \partial B) = 0$, we have

$$B = \frac{1}{\beta}(\lambda_1 AW_a - \lambda_1 AW_b + \beta PB'). \tag{11}$$

3) Update B′: The objective function regarding to B' can be written as $\min_{B'} \Phi(B') = \|X' - D'B'\|_F^2 + \alpha \|B' - V'X'\|_F^2 + \beta \|B - PB'\|_F^2$. Setting $(\partial \Phi / \partial B') = 0$, we have

$$\begin{aligned} B' = (D'^\top D' + \beta P^\top P + \alpha I)^{-1} \\ \times (D'^\top X' + \alpha V'X' + \beta P^\top B). \end{aligned} \tag{12}$$

4) Update P: The objective function turns to the following form when keeping the terms relevant only to P: $\min_P \Phi(P) = \lambda_2 \|P\|_F^2 + \beta \|B - PB'\|_F^2$. Let $(\partial \Phi / \partial P) = 0$, the closed-form solution of P is

$$P = \beta BB'^\top (\beta B'B'^\top + \lambda_2 I)^{-1}. \tag{13}$$

5) Update V and V′: The objective function reduces to $\min_V \Phi(V) = \|A - VX\|_F^2$, when removing all the terms irrelevant to V. Setting $(\partial \Phi / \partial V) = 0$, we have

$$V = AX^\top (XX^\top + \theta I)^{-1} \tag{14}$$

where $\theta = 10^{-3}$ is a small regularization parameter. We can update V' in the similar way.

6) Update D and D′: Keeping the terms relevant only to D, the objective function becomes

$$\min_D \Phi(D) = \|X - DA\|_F^2 \text{ s.t. } \|d_i\| \leq 1, \quad i = 1, \ldots, k. \tag{15}$$

The famous alternating direction method of multipliers (ADMM) algorithm can be employed to effectively solve this problem [44]. Similar solution to D' can be obtained.

Algorithm 1 Optimization of DSPDL

Input: Training data X and X', parameters λ_1 and λ_2.
1. Initialize A, B, B', P, V, V' D, and D'.
while not converged **do**
2. Fix other variables and update A according to (10);
3. Fix other variables and update B according to (11);
4. Fix other variables and update B' according to (12);
5. Fix other variables and update P according to (13);
6. Fix other variables and update V and V' using (14);
7. Fix other variables and update D and D' using ADMM algorithm.
end while
Output: D, D' and P.

The above-mentioned procedures are repeated until convergence. Algorithm 1 outlines the optimization process.

B. Convergence and Complexity Analysis

Our DSDPL model is derived from the projective dictionary learning model proposed in [44], in which the model convergence has been well studied. Similarly, we divide all the variables to be optimized into three groups, i.e., $\{A, B'\}$, $\{D, D', V, V', P\}$, and $\{B\}$, where the variables in each group are separable during the model optimization and can be treated as one during the model optimization. Meanwhile, the objective function is convex with respect to each group when the others are fixed. In this way, we formulate the optimization of our model as a multiconvex optimization problem, the convergence of which has been studied in [49] and [50]. We will also show the convergence of the proposed model through empirical study in the experimental part.

In the training phase, $A, B, B', P, V, \ V', D$, and D' are updated alternatively. The cost of updating A in each iteration is $O(k^3 + kdn + kn^2)$, that of updating B is $O(kn^2 + k^2n)$, that of updating B' is $O(k^3 + kdn + k^2n)$, that of updating P is $O(k^3 + k^2n)$, that of updating V and V' is $O(d^3 + kdn)$, and that of updating D and D' is $O(\tau(kdn + k^3 + k^2d + d^2k))$, where τ is the iteration number in ADMM algorithm for updating D and D'.

C. Model Comparison

In this section, we compare our DSPDL model with the three most relevant models to highlight our novelty: semi-coupled dictionary learning (SCDL) [51], CPDL [24], and SLD^{2}L [15].

SCDL is developed for photo-sketch synthesis and image superresolution. It requires high time consumption to solve the sparse coding problem, while our proposed DSPDL model can be solved efficiently due to the adoption of the projective dictionary learning technique. In addition, SCDL is developed to uncover the relationship between different image styles of the same instance, so that it essentially neglects the discriminative information among instances. In contrast, DSPDL is designed for ReID, we incorporate discriminative information to learn dictionaries which can help to distinguish images of the same identities from those of different ones. CPDL is designed for ReID but it neglects the fact that great image resolution divergences could comprise the generalization ability of the learned dictionaries, when directly pushing close the new codings of images of the same person. Moreover, similar to SCDL, CPDL does not incorporate interperson dissimilarity to enhance the discriminative power of the dictionaries.

SLD^{2}L is more closely related to DSPDL: both learn semi-coupled dictionaries that are robust with resolution changes. However, DSPDL differs from SLD^{2}L in the following aspects: First, we adopt the more efficient, also more powerful, projective dictionary technique; while SLD^{2}L uses the traditional dictionary learning technique. Second, SLD^{2}L segments images into small patches, clusters the patches into groups, and learns a set of dictionary pairs for all corresponding LR and HR image patch clusters. Due to the clusterwise dictionary learning strategy, SLD^{2}L is extremely complicated: It comprises of 15 terms, 9 parameters, and dozens of variables. Solving such a complicated model is definitively a time-consuming task. It is also hard to balance all the terms and tune the parameters to the state that is robust in various scenarios. This is why the parameters for SLD^{2}L are set data set by data set in the experiments. Different from SLD^{2}L, DSPDL learns only one pair of dictionaries from all images so that our model is much simpler: we have only five terms and two parameters. Therefore, our model can be easily and efficiently solved, and promise stable performances on different data sets with fixed parameters. Third, we incorporate positive and negative pair information in a parameterless graph embedding fashion, but SLD^{2}L simply combines several separated terms, the weights of which are hard to balance.

IV. DSDPL for LR-ReID

In this section, we present the application of the proposed DSPDL model for the LR-ReID problem, first for the uniform resolution gap scenario and then the variational resolution gap scenario.

A. Uniform Resolution Gap

Denoted by $L = [X_l, T_l] \in \mathbb{R}^{d\times(n+m)}$ and $H = [X_h, T_h] \in \mathbb{R}^{d\times(n+m')}$ two pedestrian image sets of LR and HR, respectively. For this scenario, it is assumed that there is a uniform resolution gap between the images from L and H. We first learn a pair of dictionaries D_l and D_h, as well as the mapping function P using the proposed DSPDL model by feeding it with training data X_l and X_h. With D_l, D_h, and P, we can obtain perform ReID as outlined in Algorithm 2. For convenience, we will later refer this algorithm as *DSPDL-Uni*.

B. Variational Resolution Gaps

The above-mentioned DSPDL-Uni algorithm assumes that there is a uniform resolution gap between the LR and HR pedestrian images, such that learning a mapping function can mitigate the gap. This assumption holds, in practice, when the images captured by each individual camera have similar resolutions, and the cross-camera resolution gap is a constant one. However, it is more practical that images captured by a camera vary a lot in resolutions and the resolution gaps vary

Algorithm 2 DSPDL for LR-ReID With Uniform Resolution Gap

Input: Training sets X_l and X_h, test sets T_l and T_h.

1. Learn D_l, D_h and P by **Algorithm 1** with X_l and X_h.
For each $l_i \in T_l$ **do**
  For each $h_j \in T_h$ **do**
    2. Calculate new coding f_l^i of l_i under D_l;
    3. Calculate new coding g_h^j of h_j under D_h;
    4. Calculate $f_h^i = Pf_l^i$;
    5. $D_{ij} = \|f_h^i - g_h^j\|_2^2$.
  end for
end for

Output: D.

person by person across different cameras. In this case, learning a single mapping function shall be incapable of modeling the nonuniform gaps. To solve this problem, we propose to use DSPDL to learn multiple pairs of dictionaries as well as multiple mapping functions, and combine the results obtained from all dictionary pairs by novel fusion technique.

Suppose we have training image sets $\mathcal{X}_l = \{X_l^s\}_{s=1}^S$ and $\mathcal{X}_h = \{X_h^s\}_{s=1}^S$ of LR and HR, respectively, with each set $X_l^s \in \mathcal{X}_l$ corresponding to set $X_h^s \in \mathcal{X}_h$. The image resolution gap of persons within any pair $(X_l^s,\ X_h^s)$ is a constant one, but this constant could be different with those of the other pairs $(X_l^k,\ X_h^k)$, $\forall k \neq s$. By feeding DSPDL with every pair $(X_l^s,\ X_h^s)$, we can learn the set $\mathcal{Q} = \{Q_s\}_{s=1}^S$, with $Q_s = (D_l^s, D_h^s, P^s)$ being the output corresponding to the sth pair. After obtaining $\mathcal{Q}$, we can evaluate the test image sets $T_l \in \mathbb{R}^{d\times m}$ and $T_h \in \mathbb{R}^{d\times m'}$, from which the resolution gaps of the images may vary person by person.

Given a probe image $l_i \in T_l$, with every $Q_s \in \mathcal{Q}$, we can calculate its distances to all gallery images by following the steps outlined in Algorithm 2, and obtain $D_e = \{d_s^j | s = 1, 2, \ldots, S;\ j = 1, 2, \cdots, m'\}$. Under each $Q_s \in \mathcal{Q}$, we can sort the distances of l_i to the gallery images T_h, $\{d_s^j | j = 1, 2, \cdots, m'\}$, and obtain a ranking list $L^s = \{r_1, r_2, \cdots, r_j, \cdots, r_{m'}\}$, where r_j indicates the jth closest gallery images relative to l_i. Collecting the ranking lists of l_i obtained based on all $Q_s \in \mathcal{Q}$, we get the ranking list set $\mathcal{L} = \{L^s\}_{s=1}^S$. With this formulation, each gallery $h_j \in T_h$ has multiple (S) distances and ranks relative to the given probe l_i. The immediate problem to be solved is how to fuse the multiple distances or ranks to reach a final one. One direct solution is to assign them with different weights, Which, however requires carefully tuning the weights data set by data set. Instead, we propose a novel fusion strategy to get the final matching results.

Inspired in [52], we utilize k-reciprocal nearest neighbors (RNNs) to rerank the initial ranking results. The k-RNNs of l_i under each $L^s \in \mathcal{L}$ is defined as

$$R^s = \{h_j | h_j \in N^s(l_i) \wedge l_i \in N^s(h_j)\} \tag{16}$$

where $N^s(l_i)$ is the set of k-nearest neighbors (NNs) of l_i from T_h, i.e., the top k elements of L^s; $N^s(h_j)$ is the set of k-NNs of h_j from T_l, which can be obtained by using h_j as the query to retrieve its matches in the probe data set T_l. k-RNNs involve cross-camera validation, i.e., a probe and its true matches should mutually be the cross-camera k-NNs of the other, thereby promoting the possibility of the true matches being retrieved in the tops. Note that our strategy for obtaining the k-RNNs is different from [52], in which probe and gallery images are mixed when searching for the neighbors. We consider only cross-camera neighbors, which fits better for ReID.

Aggregating the k-RNN sets of l_j derived from all ranking lists from $\mathcal{L}$, we get $\mathcal{R} = \{R^s\}_{s=1}^S$. It is expected that there are considerable overlaps among the k-RNN sets, because the same gallery image can be highly ranked under dictionaries learned with different resolution gaps. In fact, the more times a gallery image is included in the k-RNN sets $\mathcal{R}$, the higher possibility it is a true match for the probe. With this consideration, we define the union of the k-RNNs sets $U = \{\cup R^s\}_{s=1}^S$ as the candidate matches for the probe. For each $u_j \in U$, we calculate its appearing frequency in $\mathcal{R}$ as

$$f(u_j) = \frac{\text{sum}(V_{u_j})}{S} \tag{17}$$

where V_{u_j} is a binary vector of length S whose sth element indicates if u_j belongs to R^s, that is,

$$V_{u_j}^s = \begin{cases} 1, & \text{if } u_s \in R^s \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

We sort elements in U according to their appearing frequencies and keep the top-k elements $\hat{U}$ as the candidates.

To further rank the candidates in $\hat{U}$, we resort to their ranks in the ranking list set $\mathcal{L}$. Let $r_{u_j}^s$ be the ranks of $u_j \in \hat{U}$ in $L^s \in \mathcal{L}$. We define the ranking distance between l_i and u_j as

$$r(u_j) = 1 - \exp\left(-\frac{\sum_{i=1}^S r_{u_j}^s}{\max\left(\left\{\sum_{i=1}^S r_{u_j}^s\right\}_{j=1}^k\right)}\right). \tag{19}$$

With this definition, we finally define the ranking distance between the probe l_i and any image h_j in the gallery T_h as

$$d_r^j = \begin{cases} r(h_j), & \text{if } h_j \in \hat{U} \\ 1, & \text{otherwise.} \end{cases} \tag{20}$$

Following [52], we combine the ranking distance and the Euclidean distances, and define the final distance between l_i and any image $h_j \in T_h$ as

$$D_r(l_i, h_j) = (1 - \kappa)d_r^j + \kappa\frac{\sum_{s=1}^S d_s^j}{S} \tag{21}$$

where $\kappa \in [0, 1]$ is the tradeoff parameter balancing the two components. One can observe that the second component of the above distance function is the averaging of the multiple Euclidean distances between l_i and h_j. We use this simple strategy of fusing multiple Euclidean distances to avoid introducing extra hyperparameters.

Algorithm 3 outlines the main steps of applying DSPDL for LR-ReID with variational resolution gaps. We will refer this algorithm as *DSPDL-Var.*

Algorithm 3 DSPDL for LR-ReID With Variational Resolution Gap

Input: Training image $\mathcal{X}_l = \{X_l^s\}_{s=1}^S$ and $\mathcal{X}_h = \{X_h^s\}_{s=1}^S$; test image T_l and T_h.

For each $X_l^s \in \mathcal{X}_l^s$ and $X_h^s \in \mathcal{X}_h$ **do**
1. Learn the triple $Q_s = (D_l^s, D_h^s, P^s)$ by feeding **Algorithm 1** with X_l^s and X_h^s;

end for
For each $l_i \in T_l$
2. Obtain the ranking lists $\mathcal{L} = \{L^s\}_{s=1}^S$ using $\mathcal{Q} = \{Q_s\}_{s=1}^S$;
3. Calculate the k-RNN sets $\mathcal{R} = \{R^s\}_{s=1}^S$ for l_i using all ranking lists from $\mathcal{L}$;
4. Obtain the candidate match set $\hat{U}$ from the union the k-RNN sets $\mathcal{R}$ based on its appearing frequency defined in (17);
5. Calculate the ranking distance between l_i and every sample from T_h based on (20);
6. Calculate the final distance between l_i and every sample from T_h as defined in (21).

end for

Output: The final distance between each sample l_i from T_l and every sample from T_h.

V. Experiments

We employ five widely used data sets for performance evaluation: VIPeR [53], CUHK01 [54], PRID450S [55], QMUL-iLIDS [56], and CUHK02 [54]. The following ReID methods are employed for comparison: metric learning-based methods, LFDA [9], PCCA [7], XQDA [10], and MLAPG [32]; feature description or learning-based methods, unsupervised salience learning [17], person re-identification by saliency matching [2], mid-level filters [54], and coupled marginalized auto-encoders (CMAE) [14]; deep learning-based methods, Deeplist [57]; dictionary learning-based methods, SLD^{2}L [15], SCDL [51], CPDL [24], and sample specific support vector machine (SSSVM) [25]. Note that SCDL is originally developed for photo synthesis and image superresolution; we adapt it to ReID by feeding it with ReID data (the same as ours) and tune the parameters carefully. Among these methods, CMAE and SLD^{2}L are specifically designed for the LR-ReID. For a fair comparison, whenever possible (i.e., the implementations are public and the used feature can be replaced), the same local maximal occurrence features are used as input [10]. Otherwise, the results generated by the defaulted feature extraction methods or the reported results on the same images are compared. We adopt the standard cumulated matching characteristics result as the evaluation metric.

The proposed DSPDL model has two major parameters λ_1 and λ_2. λ_1 balances the graph regularizer and the data fidelity terms in the objective function. We will analyze it later. λ_2 controls the scale of a variable, such that a small value is preferred. We empirically set it as $\lambda_2 = 0.01$. When applying the proposed DSPDL model for the variational resolution gap scenario, i.e., DSPDL-Var, two extra parameters are introduced, k and κ. k is the number of neighbors collected for each image when calculating the ranking distance, we empirically set it as 20 for all experiments κ is used to balance the two types of distances; its impacts on the performance of DSDPL-Var will also be analyzed later.

A. Comparative Results

The uniform resolution gap assumption between the probe and gallery image sets is commonly adopted by existing LR-ReID methods. The proposed DSPDL-Uni also underlies this assumption, but the proposed DSPDL-Var is less dependent on it. For a fair comparison, we conduct experiments for both the uniform and variational resolution gap scenarios.

1) Uniform Resolution Gap: We follow the experimental setup of existing LR-ReID methods [12], [14], [15]. Given the probe and gallery images X_p and X_g, we downsample X_p with the rate 1/8 and keep X_g unchanged to simulate the large and uniform resolution gap. The proposed DSPDL-Uni method can directly be applied to this experimental setting. To test the effectiveness of the designed graph regularization term, we also apply the SPDL model on the LR-ReID problem, following the same steps as that for DSPDL-Uni. Analogously, we refer it as SPDL-Uni.

DSPDL-Var cannot directly be applied to this setting, because it requires training data of variational resolution gaps. However, we can make a slight change on the experimental data to evaluate it. Instead of keeping the gallery images X_g unchanged, we downsample the images with the rate 1/2, 1/4, and 1/8, generating X_g^2, X_g^4, and X_g^8, respectively. In this way, probe image set X_p and can be paired with every gallery image set from $\{X_g, X_g^2, X_g^4, X_g^8\}$, and each pair can be fed to DSPDL-Uni to get a ReID result. We then employ DSPDL-Var to rerank and fuse all the ReID results and get a final one.

a) VIPeR: The VIPeR data set is one of the most widely used data sets for ReID. It contains 632 persons with each having a pair of images. All the images are normalized to 128×48 pixels. There are significant viewpoint changes, pose variations, and illumination differences across the cameras. The synthesized great resolution differences make it even harder to match the images of the same identifies. By randomly dividing the data set into training and testing parts of equal size, i.e., 316 image pairs for training and the other 316 pairs for testing, and repeating the randomly division procedure for 10 times, we obtain the average matching rates.

Table I shows the top rank 1, 5, 10, and 20 matching rates of our proposed methods (SPDL-Uni, DSPDL-Uni, and DSPDL-Var), and the competing methods on this data set. We can observe that DSPDL-Uni beats SPDL-Uni by 2% or 3% for all the ranks, substantiating the effectiveness of the designed graph regularization term on boosting the discriminability of the learned dictionaries. It is also observed that DSPDL-Var outperforms DSPDL-Uni, evidencing the effectiveness of the reranking and fusion strategy.

On the other hand, compared with the competing methods, our methods exhibit advantages. Compared with the best feature learning-based method CMAE [14], DSPDL-Uni gains about 2.5% and 11% improvements for the rank-1 and rank-5 matching rates, respectively. Compared with the metric

TABLE I

TOP r MATCHING RATES (%) ON THE VIPeR DATA SET. THE BEST/SECOND BEST RESULTS ARE MARKED IN RED/BLUE

Methods	$r = 1$	$r = 5$	$r = 10$	$r = 20$
USL [17]	14.87	36.08	44.30	56.96
PRSM [2]	16.20	34.24	45.06	56.96
MLF [54]	16.65	32.91	44.87	57.91
CAE [14]	25.95	50.00	64.37	79.75
LFDA [9]	9.57	28.80	43.33	60.94
PCCA [7]	8.55	27.39	41.17	58.68
XQDA [10]	23.26	53.86	70.03	84.68
MLAPG [32]	24.72	54.91	69.62	83.54
CPDL [24]	19.02	49.97	67.12	81.49
SCDL [51]	22.56	54.48	69.59	84.21
SLD^{2}L [15]	16.86	41.22	58.06	79.00
SSSVM [25]	24.53	53.04	65.54	78.89
Deeplist [57]	25.95	58.23	70.57	83.54
SPDL-Uni	26.27	58.86	73.73	85.45
DSPDL-Uni	28.51	61.08	76.11	88.13
DSPDL-Var	30.38	63.61	76.27	87.66

learning-based methods, DSPDL-Uni achieves about 4% and 6% gains in the rank-1 and rank-5 matching rates, respectively, over MLAPG, the best method in this category. Our method is based on dictionary learning, and DSPDL-Uni gains the rank-1 and rank-10 matching rate promotions of near 4% and 10.5%, respectively, over the best existing dictionary learning-based method SSSVM [25]. Recent years have witnessed the overwhelming advantages of deep learning on various research domains, such as image classification, object detection, and so on, partially owing to the richness of labeled data. The VIPeR data set is relatively small, so it is hard to train a complex and powerful deep model utilizing the limited labeled data. This explains the inferior results of the recent deep learning-based ReID method, Deeplist [57], to that of DSPDL-Uni: It reaches 2.5% and 5.5% performance gains for the rank-1 and rank-10 matching rates, respectively, over Deeplist. The performance margins are even more remarkable for DSPDL-Var over Deeplist.

We can find that though CMAE and SLD^{2}L are designed specifically for matching pedestrian images of great resolution divergences, they surprisingly perform worse than methods which do not target at this degenerated scenario. For example, although CMAE has a small advantage over all the other competing methods for the rank-1 matching rate, its rank-5 matching rate is much lower than those of XQDA, MLAPG, SCDL, and SSSVM. The superresolution-based ReID method SLD^{2}L performs even worse. However, our proposed SPDL-Uni and DSPDL-Uni do perform better than all the competing methods, and our advantages are significant in many cases.

b) CUHK01: The CUHK01 data set contains 3884 pedestrian images of 971 persons in two camera views, where each person has two images in each view. We take the first one in each view for the experiment. Images in this data set are of HR, which could be a beneficial factor for ReID. By randomly selecting half identities (485 persons) for training and the other half (486 persons) for testing, and repeating the trials for 10 times, we obtain the matching rates shown in Table II. Similar as the observations on the VIPeR data set, our methods beat all the competing methods, showing our the advantages in handling great resolution gaps. Meanwhile, DSPDL-Var gains better results than DSPDL-Uni, which, on the other hand, outperforms SPDL-Uni. This again evidences the effectiveness of the proposed fusion and reranking strategy and our designed graph regularization term.

c) PRID450S: The PRID450S data set is based on the PRID 2011 data set [55], including 450 image pairs from two cameras. The partial occlusions and viewpoint changes make it a challenging data set for ReID. The synthetic great resolution gap makes it even harder for LR-ReID. Table II shows the average matching rates over 10 trials of evenly dividing the data set into training and test partitions. Similar as what we observed from the other data sets, the proposed DSPDL-Uni beats all the competing methods and the DSPDL-Var achieves even higher matching accuracies, especially for the rank-1 and 10 matching rates, where DSPDL-Var gains about 3.5% improvement relative to DSPDL-Uni. This once again proves the effectiveness of the proposed reranking and fusion technique.

d) QMUL-iLIDS: The QMUL-iLIDS data set [56] consists of 476 images of 119 identities; each person has four images on average. We randomly select two images for each person and downsample one of them at the rate 1/8, and keep the other unchanged to simulate the resolution difference. Among the 119 images pairs, we randomly select 59 pairs for training, and the left 60 image pairs for testing. We run the experiment 10 times and calculate the average matching rates. The rank-1, 5, 10, and 20 matching rates are given in Table II. We can see that all the methods achieve high matching rates quickly as the rank increases. This is because this data set is small such that for each probe image, there are only 60 images to be queried. We can also observe that our proposed methods beat all the competing methods by large margins, except a slight inferior to SSSVM for the rank-20 matching rate. It is worthy of noting that the advantages of our proposed methods over existing ones on this data set are more significant than those on the other data sets. This indicates that our methods have some inclinations to smaller data sets, and more readily beats existing methods in this preferred scenario.

e) CUHK02: The CUHK02 data set is larger than all the above four data sets. It contains 7264 images of 1816 persons; two images for each person from each camera view. The CUHK02 data set is an extension of the CUHK01 data set, so we follow the same experimental setting. Specifically, we take the 908 persons for training and the rest 908 persons for testing. For each person, we choose two images for experiments. The experimental results are given in Table II. Similar observations as we get in the other smaller data sets that DSPDL-Var and DSPDL-Uni obtain the best performance.

2) Variational Resolution Gaps: The above-mentioned experiments show that our proposed methods outperform the state-of-the-art ones for handling the uniform resolution gap between probe and gallery sets. In this section, we conduct experiments to verify the effectiveness of the proposed DSPDL-Var for the harder yet more practical scenario where the resolution gaps between probe and gallery image sets vary person by person. We employ the same data sets and adopt

TABLE II

TOP r MATCHING RATES (%) ON FIVE DATA SET WITH VARIATIONAL RESOLUTION GAPS. THE BEST RESULTS ARE IN BOLD

	CUHK01				PRID450S				QMUL-iLIDS				CUHK02			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
SCDL [51]	14.05	33.37	44.57	56.40	35.6	65.33	76.44	87.47	45.33	79.17	86.50	94.50	18.94	36.57	45.10	50.58
CPDL [24]	16.50	37.20	48.46	61.42	28.31	58.84	73.47	85.11	42.33	71.50	83.67	92.83	16.60	39.00	50.95	58.03
XQDA [10]	18.97	43.58	55.80	68.19	37.24	66.89	78.09	88.09	50.67	79.17	86.00	93.33	17.6	39.69	51.35	58.41
SSSVM [25]	17.02	37.90	48.02	59.71	37.56	66.31	77.02	86.04	41.33	75.67	88.17	96.67	19.19	39.94	50.53	56.96
MLAPG [32]	19.51	40.41	52.47	65.16	37.73	65.78	77.64	87.87	46.67	77.17	85.67	93.33	19.70	40.70	50.48	57.14
SPDL-Uni	20.66	44.84	56.79	69.22	39.11	65.33	77.33	88.44	52.83	79.00	87.50	95.17	17.29	36.01	48.13	53.19
DSPDL-Uni	21.75	46.50	58.27	69.57	39.64	68.18	77.78	87.82	55.17	82.00	90.67	95.67	19.49	38.11	51.10	57.27
DSPDL-Var	24.69	46.09	59.67	69.34	43.11	69.42	81.07	90.67	61.67	83.00	90.67	95.00	23.68	42.84	50.55	60.35

TABLE III

TOP r MATCHING RATES (%) ON FIVE DATA SET WITH VARIATIONAL RESOLUTION GAPS. THE BEST RESULTS ARE IN BOLD

	VIPeR			CUHK01			PRID450S			QMUL-iLIDS			CUHK02		
	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$
SCDL [51]	18.67	44.94	63.92	11.11	24.49	34.57	31.11	61.29	73.02	46.00	73.17	85.83	5.80	14.65	20.53
CPDL [24]	17.94	44.68	60.32	15.93	37.12	48.15	30.58	59.51	70.98	41.67	75.00	85.00	14.72	35.56	47.44
XQDA [10]	19.59	45.03	61.20	15.66	37.90	49.20	35.29	65.20	76.93	48.33	76.67	81.67	15.56	35.40	47.33
SSSVM [25]	21.20	46.27	60.32	16.69	36.48	46.91	37.51	65.16	75.28	43.33	78.67	89.33	15.72	33.28	43.24
MLAPG [32]	21.36	48.83	63.29	17.90	37.45	47.74	36.76	65.02	76.22	43.33	76.67	88.33	15.64	35.79	47.36
DSPDL-Uni	23.42	46.20	59.49	16.87	36.42	45.47	36.53	67.16	78.67	50.00	75.50	88.83	15.75	34.80	44.49
DSPDL-Var	25.63	49.05	65.82	19.96	40.33	50.21	39.56	67.56	78.22	57.67	78.33	88.33	17.40	37.44	45.48

the same protocols as we conduct experiments for the uniform resolution gap case. The only difference lies in the way we prepare the experimental data. Given the images X_p and X_g from two cameras, for the uniform resolution gap scenario, we downsample all images from one camera X_p at the rate of 1/8, and keep the images from X_g unchanged to simulate the uniform resolution gap. Here, while we downsample X_p at the rate of 1/8 as well, but we evenly divide X_g into four parts X_g^0, X_g^1, X_g^2, and X_g^3. We keep X_g^0 unchanged, while downsampling X_g^1, X_g^2, and X_g^3 by the rates of 1/2, 1/4, and 1/8, respectively. We assemble X_g^0, X_g^1, X_g^2, and X_g^3 as the new gallery set for experiments. Note that in the training stage of the proposed DSPDL-Var, we need to know the resolution gaps between probe and gallery image sets and learn the pair of dictionaries and mapping function corresponding to each resolution gap. However, in the testing stage, the resolution gap between a probe image and any gallery image is unknown. This is practical because we can train our model with cross-view images of known resolution gaps, and use the model to evaluate cross-view image of unknown resolution gaps.

Several most competitive competing methods are employed for experiments. The comparative results are given in Table III. From the table, we can observe that the proposed DSPDL-Var performs the best on all the five data sets, with an exception for the rank-5 and rank-10 matching accuracies on the QMUL-iLIDS data set, where DSPDL-Var gets slightly worse results than SSSVM. However, the interesting thing is that, in the same data set, DSPDL-Var beats SSSVM by more than 14% for the rank-1 matching accuracy. We speculate the reason for this inconsistency is that the QMUL-iLIDS data set is quite small and a fraction of the cross-view person images are easy to be matched, while the others are hard. The proposed DSPDL-Var is much more effective than the competing methods for handling those easy ones and return the correct matches in the most top ranks. However, some of those hard ones are too intractable such that all methods cannot well handle them, and their correct matches cannot be returned even we enlarge the ranking list. On the other hand, as we enlarge the ranking list, the competing methods catch up DSPDL-Var and return the correct matches for those easy ones. This explains the great advantage of DSPDL-Var for the rank-1 matching rate, while being only comparable to some of the competing methods for the higher ranks. We also find that though DSPDL-Uni is not specifically designed for variational resolution gaps, it remains competitive relative to the competing methods.

B. Further Analysis

1) Parameter Analysis: Our proposed DSPDL model has an important parameter λ_1, which balances the graph regularization term and the data fidelity terms in the objective function. We vary its value from 10^{-3} to 10, and compute the rank-1 matching rates of DSPDL-Uni on all the five data sets. The results are shown in Fig. 2(a). We observe that DSPDL reaches the best performances when $\lambda_1 = 0.1$. Therefore, we adopt this setting as default in our method

One extra important parameter $\kappa \in [0, 1]$ is introduced when applying the DSPDL model for the variational resolution gap scenario. κ balances the Euclidean distance and the ranking distance in the final distance function. When $\kappa = 0$, only the ranking distance is considered, while only Euclidean distance is considered when $\kappa = 1$. Fig. 2(b) shows the change of the matching rate with respect to κ on the VIPeR data set. We can observe from the figure that a combination of both types of distances does help to boost the ReID performance.

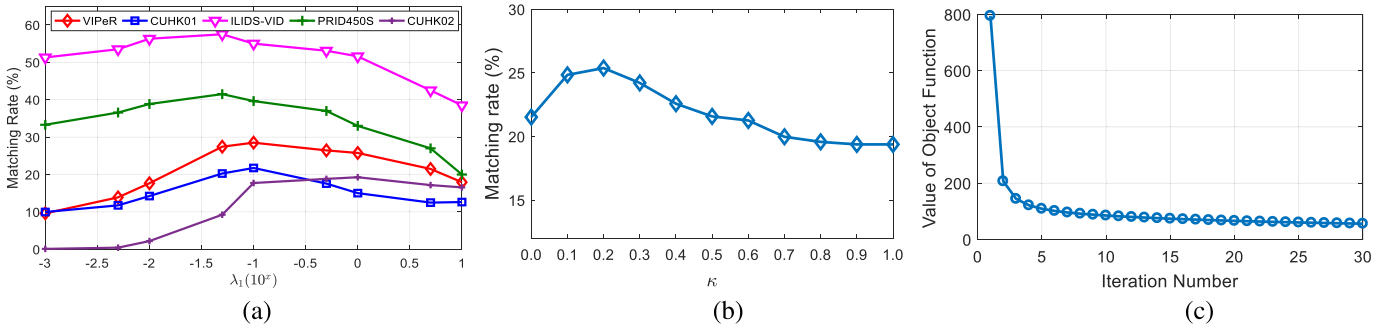


Fig. 2. (a) Rank-1 matching rates of the proposed DSPDL-Uni on the five data sets with different values of λ_1. (b) Change of the Rank-1 matching rate of DSPDL-Var with respect to κ on the VIPeR data set. (c) Convergence curve of the proposed DSPDL model on the VIPeR data set.

TABLE IV
RUNNING TIME ON THE VIPeR DATA SET FOR ONE TRIAL

Methods	SLD^{2}L	SCDL	DSPDL
time (s)	∼1200.00	69.04	22.63

2) Convergence Analysis: Fig. 2(c) shows the changes of the value of the objective function of DSPDL with the increase in iteration times on the VIPeR data set. We can see that with a small number of iterations, our objective function turns to be stable, which shows the good convergence property of our model.

3) Running Time: The proposed DSPDL model is based on the projective dictionary learning [44], which avoids solving the inefficient sparse coding problem as traditional dictionary learning methods do. Therefore, it is expected to take less time to train the DSPDL model than that of traditional dictionary learning-based ReID models. To verify this, we compare the training time of DSPDL and another two dictionary learning-based methods, SCDL and SLD^2L. The result in Table IV shows that the proposed DSPDL takes much less training time than the other dictionary learning-based methods.

VI. CONCLUSION

We presented a new DSPDL model and applied it to the LR person ReID problem. DSPDL adopted the efficient projective dictionary learning technique and learned a mapping function along with a pair of dictionaries to model the correspondence of the cross-view data. A parameterless cross-view graph regularizer was designed to incorporate both positive and negative pair information, so that the discriminability of the dictionaries is enhanced. To extend DSPDL for the scenario where there are variational resolution gaps between cross-camera pedestrian images, we proposed to use DSPDL to learn multiple pairs of dictionaries and multiple mapping functions, and formulated a novel technique to fuse and reranking the ReID results obtained from all dictionary pairs. Experimental results on five data sets showed our method outperforms the state of the art, often by large margins, for both the uniform resolution gap scenario and the variational resolution gap scenario.

REFERENCES

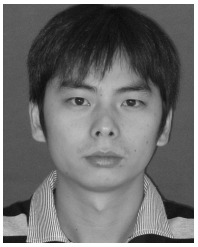
[1] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by Fisher vectors for person re-identification," in *Proc. ECCV Workshops Demonstrations*, 2012, pp. 413–422.
[2] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proc. ICCV*, 2013, pp. 2528–2535.
[3] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang, "Similarity learning on an explicit polynomial kernel feature map for person re-identification," in *Proc. CVPR*, 2015, pp. 1565–1573.
[4] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. CVPR*, 2015, pp. 3908–3916.
[5] 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.
[6] W.-S. Zheng, S. Gong, and T. Xiang, "Transfer re-identification: From person to set-based verification," in *Proc. CVPR*, 2012, pp. 2650–2657.
[7] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. CVPR*, 2012, pp. 2666–2672.
[8] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. CVPR*, 2013, pp. 3318–3325.
[9] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *Proc. ECCV*, 2014, pp. 1–16.
[10] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. CVPR*, 2015, pp. 2197–2206.
[11] K. Li, Z. Ding, K. Li, Y. Zhang, and Y. Fu. (2018). "Support neighbor loss for person re-identification." [Online]. Available: https://arxiv.org/abs/1808.06030
[12] X. Li, W.-S. Zheng, X. Wang, T. Xiang, and S. Gong, "Multi-scale learning for low-resolution person re-identification," in *Proc. CVPR*, 2015, pp. 3765–3773.
[13] Z. Wang, R. Hu, Y. Yu, J. Jiang, C. Liang, and J. Wang, "Scale-adaptive low-resolution person re-identification via learning a discriminating surface," in *Proc. IJCAI*, 2016, pp. 2669–2675.
[14] S. Wang, Z. Ding, and Y. Fu, "Coupled marginalized auto-encoders for cross-domain multi-view learning," in *Proc. IJCAI*, 2016, pp. 2125–2131.
[15] X.-Y. Jing *et al.*, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," in *Proc. CVPR*, 2015, pp. 695–704.
[16] K. Li, Z. Ding, S. Li, and Y. Fu, "Discriminative semi-coupled projective dictionary learning for low-resolution person re-identification," in *Proc. AAAI*, 2018, pp. 2331–2338.
[17] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. CVPR*, 2013, pp. 3586–3593.
[18] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. CVPR*, 2013, pp. 3610–3617.
[19] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *Proc. CVPR*, 2016, pp. 1268–1277.

[20] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, and J. Wang, "Person re-identification with correspondence structure learning," in *Proc. ICCV*, 2015, pp. 3200–3208.
[21] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proc. CVPR*, 2016, pp. 1363–1372.
[22] L. An, X. Chen, S. Yang, and X. Li, "Person re-identification by multi-hypergraph fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2763–2774, Nov. 2017.
[23] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Person re-identification by unsupervised l_1 graph learning," in *Proc. ECCV*, 2016, pp. 178–195.
[24] S. Li, M. Shao, and Y. Fu, "Cross-view projective dictionary learning for person re-identification," in *Proc. IJCAI*, 2015, pp. 2155–2161.
[25] Y. Zhang, B. Li, H. Lu, A. Irie, and X. Ruan, "Sample-specific SVM learning for person re-identification," in *Proc. CVPR*, 2016, pp. 1278–1287.
[26] X. Liu, M. Song, Q. Zhao, D. Tao, C. Chen, and J. Bu, "Attribute-restricted latent topic model for person re-identification," *Pattern Recognit.*, vol. 45, no. 12, pp. 4204–4213, 2012.
[27] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao, "Multi-task learning with low rank attribute embedding for person re-identification," in *Proc. ICCV*, 2015, pp. 3739–3747.
[28] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, and Y. Yang. (2017). "Improving person re-identification by attribute and identity learning." [Online]. Available: https://arxiv.org/abs/1703.07220
[29] A. Schumann and R. Stiefelhagen, "Person re-identification by deep learning attribute-complementary information," in *Proc. CVPR*, 2017, pp. 1435–1443.
[30] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. CVPR*, 2012, pp. 2288–2295.
[31] X. Xu, W. Li, and D. Xu, "Distance metric learning using privileged information for face verification and person re-identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3150–3162, Dec. 2015.
[32] S. Liao and S. Z. Li, "Efficient PSD constrained asymmetric metric learning for person re-identification," in *Proc. ICCV*, 2015, pp. 3685–3693.
[33] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. CVPR*, 2015, pp. 1846–1855.
[34] S. Li, M. Shao, and Y. Fu, "Person re-identification by cross-view multi-level dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.
[35] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proc. ICPR*, 2014, pp. 34–39.
[36] 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. CVPR*, 2016, pp. 1335–1344.
[37] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. CVPR*, 2014, pp. 152–159.
[38] L. Wu, C. Shen, and A. van den Hengel. (2016). "PersonNet: Person re-identification with deep convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1601.07255
[39] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Proc. CVPR*, 2016, pp. 1249–1258.
[40] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Deep attributes driven multi-camera person re-identification," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Oct. 2016, pp. 475–491.
[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE CVPR Workshop*, Jun. 2009, pp. 248–255.
[42] K.-K. Huang, D.-Q. Dai, C.-X. Ren, and Z.-R. Lai, "Learning kernel extended dictionary for face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1082–1094, May 2017.
[43] Z. Li, Z. Lai, Y. Xu, J. Yang, and D. Zhang, "A locality-constrained and label embedding dictionary learning algorithm for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 278–293, Feb. 2015.
[44] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proc. NIPS*, 2014, pp. 793–801.
[45] M. Yang, W. Liu, W. Luo, and L. Shen, "Analysis-synthesis dictionary learning for universality-particularity representation based classification," in *Proc. AAAI*, 2016, pp. 2251–2257.
[46] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[47] D. Cai, X. He, and J. Han, "Sparse projections over graph," in *Proc. AAAI*, 2008, pp. 610–615.
[48] K. Li, S. Li, Z. Ding, W. Zhang, and Y. Fu, "Latent discriminant subspace representations for multi-view outlier detection," in *Proc. AAAI*, 2018, pp. 3522–3529.
[49] H. Al-Shatri, X. Li, R. S. Ganesan, A. Klein, and T. Weber, "Maximizing the sum rate in cellular networks using multiconvex optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3199–3211, May 2016.
[50] H. Zhao, O. Stretcu, A. Smola, and G. Gordon. (2017). "Efficient multitask feature and relationship learning," [Online]. Available: https://arxiv.org/abs/1702.04423
[51] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. CVPR*, 2012, pp. 2216–2223.
[52] Z. Zhong, L. Zheng, D. Cao, and S. Li. (2017). "Re-ranking person re-identification with k-reciprocal encoding." [Online]. Available: https://arxiv.org/abs/1701.08398
[53] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, 2008, pp. 262–275.
[54] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. CVPR*, 2014, pp. 144–151.
[55] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznai, and H. Bischof, "Mahalanobis distance learning for person re-identification," in *Person Re-Identification* (Advances in Computer Vision and Pattern Recognition), S. Gong, M. Cristani, S. Yan, and C. C. Loy, Eds. London, U.K.: Springer, 2014, pp. 247–267.
[56] J. M. Berry, *Lobbying for the People: The Political Behavior of Public Interest Groups*. Princeton, NJ, USA: Princeton Univ. Press, 2015.
[57] J. Wang, Z. Wang, C. Gao, N. Sang, and R. Huang, "DeepList: Learning deep features with adaptive listwise constraint for person reidentification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 513–524, Mar. 2016.

Kai Li (S'15) received the B.Eng. and M.Eng. degrees in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA.

His current research interests include representation learning and its applications.

Mr. Li is an AAAI Student Member. He has served as a reviewer for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and so on.



Zhengming Ding (S'14–M'18) received the B.Eng. degree in information security and the M.Eng. degree in computer software and theory from the University of Electronic Science and Technology of China, Chengdu, China, in 2010 and 2013, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA, in 2018.

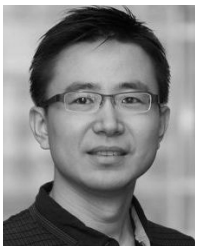
Since 2018, he has been a Faculty Member with the Department of Computer, Information and Technology, Indiana University—Purdue University Indianapolis, Indianapolis, IN, USA. His current research interests include machine learning and computer vision, especially in developing scalable algorithms for challenging problems in transfer learning and deep learning scenario.

Dr. Ding was a recipient of the National Institute of Justice Fellowship from 2016 to 2018, the Best Paper Award from SPIE in 2016, and the Best Paper Candidate from ACM MM in 2017. He is currently an Associate Editor of the *Journal of Electronic Imaging*.

Sheng Li (S'11–M'17) received the B.Eng. degree in computer science and engineering and the M.Eng. degree in information security from the Nanjing University of Posts and Telecommunications, Nanjing, China, and the Ph.D. degree in computer engineering from Northeastern University, Boston, MA, USA, in 2010, 2012, and 2017, respectively.

From 2017 to 2018, he was a Research Scientist with Adobe Research, San Jose, CA, USA. Since 2018, he has been a Tenure-Track Assistant Professor with the Department of Computer Science, University of Georgia, Athens, GA, USA. He has authored or co-authored over 65 papers at leading conferences and journals. His current research interests include robust machine learning, representation learning, visual intelligence, and behavior modeling.

Dr. Li was a recipient of the Best Paper Awards (or nominations) at SDM 2014, the IEEE ICME 2014, and the IEEE FG 2013. He serves as an Associate Editor for the *IEEE Computational Intelligence Magazine*, *Neurcomputing*, *IET Image Processing*, and *Journal of Electronic Imaging*. He has also served as a reviewer for several IEEE TRANSACTIONS and a Program Committee Member for NIPS, IJCAI, AAAI, and KDD.



Yun Fu (S'07–M'08–SM'11) received the B.Eng. degree in information engineering and the M.Eng. degree in pattern recognition and intelligence systems from Xi'an Jiaotong University, Xi'an, China, and the M.S. degree in statistics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Champaign–Urbana, IL, USA.

Since 2012, he has been an Interdisciplinary Faculty Member with the College of Engineering and the College of Computer and Information Science, Northeastern University, Boston, MA, USA. He has authored or co-authored leading journals, books/book chapters, and international conferences/workshops. His current research interests include machine learning, computational intelligence, big data mining, computer vision, pattern recognition, and cyber-physical systems.

Dr. Fu is a Fellow of IAPR, OSA, and SPIE, a Lifetime Senior Member of ACM, a Lifetime Member of AAAI and Institute of Mathematical Statistics, a Member of ACM Future of Computing Academy, Global Young Academy, AAAS, INNS, Beckman Graduate Fellow from 2007 to 2008. He serves as the Chair, a PC Member, a reviewer, and an Associate Editor of many top journals and international conferences/workshops. He was a recipient of the Seven Prestigious Young Investigator Awards from NAE, ONR, ARO, IEEE, INNS, UIUC, and Grainger Foundation, nine Best Paper Awards from IEEE, IAPR, SPIE, and SIAM, and many major Industrial Research Awards from Google, Samsung, and Adobe. He is currently an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEANING SYSTEMS.