Dependence-Aware Feature Coding for Person Re-Identification

Xiaobo Wang, Zhen Lei, *Senior Member, IEEE*, Shengcai Liao, *Senior Member, IEEE*, Xiaojie Guo, Yang Yang, and Stan Z. Li, *Fellow, IEEE*

***Abstract*—In this letter, we focus on how to boost the performance of person re-identification by exploring the discriminative information among person pairs. A novel dependence-aware feature coding framework is proposed for this task. Specifically, we employ the Hilbert–Schmidt independence criterion as the discriminative term, which is to explore the dependence between different kinds of person pairs, i.e., the same person pairs should be dependence maximized, while the different ones should be dependence minimized. Theoretical discussion and analysis on the convexity of the proposed constraint, as well as the convergence of our algorithm, are provided. Experimental results on two benchmark datasets have demonstrated the advantages of our method over the state-of-the-art alternatives.**

***Index Terms*—Feature coding, Person re-identification.**

I. Introduction

PERSON re-identification is the problem of matching people across several disjoint camera views, which has recently attracted much attention due to its potential applications such as forensic search [1], long-term multicamera tracking [2], and crowd movements analysis in public places [3]. To address this task, a commonly used pipeline is first to extract the appearance-based person representation [4]–[7], and then a metric is employed for matching them [8]–[12]. In practice, due to large viewpoint changes, illumination, different poses, background clutter, and occlusions, there is often large intraclass appearance variations, which make the extracted representations unstable. For instance, the descriptive features extracted in KISSME [13], the symmetry-driven accumulation of local features [14], color invariants [15], salient color names based descriptors [16]–[18], mid-level filters [19], and fusion of color models [20], are hard to describe the transitions among different camera views and are often with less discriminative power.

Manuscript received October 26, 2017; revised February 4, 2018; accepted February 4, 2018. Date of publication; date of current version. This work was supported in part by the National Key Research and Development Plan under Grant 2016YFC0801002, in part by the Chinese National Natural Science Foundation under Project 61473291, Project 61572501, Project 61502491, Project 61572536, and Project 61672521, and in part by AuthenMetric R&D Funds. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Marco Felipe Duarte. *(Corresponding author: Zhen Lei.)*

X. Wang, Z. Lei, S. Liao, Y. Yang, and S. Z. Li are with CBSR&NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: xiaobo.wang@nlpr.ia.ac.cn; zlei@nlpr.ia.ac.cn; scliao@nlpr.ia.ac.cn; yang.yang@nlpr.ia.ac.cn; szli@nlpr.ia.ac.cn).

X. Guo is with the School of Computer Software, Tianjin University, Tianjin 300350, China (e-mail: xj.max.guo@gmail.com).

Digital Object Identifier 10.1109/LSP.2018.2803776

Recently, to reduce the intraclass variations in feature space, various coding methods to encode the image-level features into high-level ones have been developed. The work [21] proposes a Soft-Assignment Coding (SAC) method, which uses all the bases to encode the image-level features. Wang *et al.* [22] used the Locality-constrained Linear Coding (LLC) method to exploit the locality between different samples and assign smaller coefficients to the bases that are farther away from each other in the Euclidean space. Huang *et al.* [23] by using the Salient Coding (SC) method enforce that the nearest code is much closer than others based on the saliency. However, these methods are unsupervised and simply employing the *k*-means to construct the dictionary to encode features may reduce the dictionary discriminability. To learn a good dictionary, Guo *et al.* [24] introduce pairwise constraints to enhance the dictionary discrimination for face verification. Gangeh *et al.* [25] propose a kernelized supervised dictionary learning for classification. However, all of them learn that a synthesis dictionary and an extra coding step are needed to obtain coding features. To address the issue, the dual Analysis Dictionary Learning (ADL) has drawn much attention recently.

The goal of ADL [26], [27] is to learn a transformation and directly obtain the high-level features. Instead of utilizing off-the-shelf transformations like Fast Fourier Transformation (FFT), Discrete Cosine Transform (DCT), etc., Gu *et al.* [28] try to enforce the class-specific dictionaries to well represent a certain class as well as to be ineffectual on the other classes. Wang *et al.* [29] aim to learn analysis subdictionaries by integrating a max-margin regularization term to enhance the discrimination of coding features. Yang *et al.* [30] enforce a linear classifier on the coding coefficient to jointly learn the dictionary pair. Guo *et al.* [31] incorporate a code consistent term and a triplet constraint-based local topology preserving term to improve the dictionary discriminability. However, all these works are designed for multiclass classification problem. It is not suitable for the weak labels in the person re-identification task [32]. Recently, Li *et al.* [33] employ the analysis dictionary for the person re-identification task. However, they only consider the positive pairs as the discriminative regularization, without considering the effect of negative pairs.

Based on the above analysis, in this letter, we aim to learn an analysis dictionary by exploiting a more powerful discriminative criterion to boost the task of person re-identification. For clarity, the main contributions are summarized as follows:

1) We propose a novel dependence-aware feature coding framework for the person re-identification task. Specifically, the proposed model employs the Hilbert–Schmidt Independence Criterion (HSIC) as the discriminative term, which is to make the same person pairs dependence

maximized, while the different ones dependence minimized.

2) Theoretical discussion and analysis on the discriminative term (i.e., the convexity of subproblems and the convergence of our algorithm) are provided.

II. Preliminary Knowledge

A. Discriminative Dictionary Learning (DDL)

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be the original image-level features of persons. Each column $\mathbf{x}_i$ is a feature vector, d is the dimensionality, and n is the total amount of data points. The core idea of DDL is to learn an optimized dictionary that can effectively represent each sample with sufficient discriminative ability. We denote $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{m \times n}$ as the coding features of $\mathbf{X}$ over the learned dictionary.

Synthesis Dictionary Learning (SDL): The SDL aims to learn a synthesis dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_m] \in \mathbb{R}^{d \times m}$ by solving the following problem:

$$\min_{\mathbf{D},\mathbf{Z}} \parallel \mathbf{X} - \mathbf{D}\mathbf{Z} \parallel_F^2 + \lambda_1 \parallel \mathbf{Z} \parallel_p + \lambda_2 \mathcal{L}(\mathbf{Z}) \tag{1}$$

where λ_1 and λ_2 are the tradeoff parameters, $||\mathbf{X} - \mathbf{D}\mathbf{Z}||_F^2$ stands for the reconstruction error of SDL model, p denotes the parameter of the ℓ_p-norm regularizer (e.g., ℓ_1-norm or ℓ_2-norm) to avoid the overfitting, and $\mathcal{L}$ denotes the discrimination term for $\mathbf{Z}$. Moreover, to avoid the scaling issue, additional constraints (e.g., $\mathbf{D}^T\mathbf{D} = \mathbf{I}$ or $||\mathbf{d}_i||_2 \leq 1$) on $\mathbf{D}$ are needed.

Analysis Dictionary Learning (ADL): As a dual analysis viewpoint of the commonly used SDL, ADL learns an analysis dictionary $\mathbf{P} = [\mathbf{p}_1; \mathbf{p}_2; \ldots; \mathbf{p}_m] \in \mathbb{R}^{m \times d}$ by

$$\min_{\mathbf{P},\mathbf{Z}} \parallel \mathbf{P}\mathbf{X} - \mathbf{Z} \parallel_F^2 + \lambda_1 \parallel \mathbf{Z} \parallel_p + \lambda_2 \mathcal{L}(\mathbf{Z}). \tag{2}$$

Similarly, constraints (e.g., $||\mathbf{P}||_F$ or $||\mathbf{p}^i||_2 \leq 1$) on $\mathbf{P}$ are employed for a well-regularized solution. The refined coding features can be directly obtained as $\mathbf{P}\mathbf{X}$.

B. Hilbert–Schmidt Independence Criterion

The HSIC is proposed in [34] to measure the (in)dependence of two random variables $\mathcal{X}$ and $\mathcal{Y}$. It has the following empirical definition.

Definition 1 (HSIC): Consider a series of n independent observations drawn from $p_{\mathbf{xy}}$, $\mathcal{Z} := \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)\} \subseteq \mathcal{X} \times \mathcal{Y}$, an empirical estimator of $HSIC(\mathcal{Z}, \mathcal{F}, \mathcal{G})$, is given by

$$HSIC(\mathcal{Z}, \mathcal{F}, \mathcal{G}) = (n-1)^{-2} \mathrm{tr}(\mathbf{K}_1 \mathbf{H} \mathbf{K}_2 \mathbf{H}) \tag{3}$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are the Gram matrices with $k_{1,ij} = k_1(\mathbf{x}_i, \mathbf{x}_j)$, $k_{2,ij} = k_2(\mathbf{y}_i, \mathbf{y}_j)$. $k_1(\mathbf{x}_i, \mathbf{x}_j)$ and $k_2(\mathbf{y}_i, \mathbf{y}_j)$ are the kernel functions defined in the kernel space $\mathcal{F}$ and $\mathcal{G}$, respectively. $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^T$ is a center matrix, which centers the Gram matrix to have zero mean.

It is important to note that according to (3), to maximize the dependence between two variables $\mathcal{X}$ and $\mathcal{Y}$, the empirical estimate of HSIC, i.e., $\mathrm{tr}(\mathbf{K}_1 \mathbf{H} \mathbf{K}_2 \mathbf{H})$, should be maximized.

III. Problem Formulation

To formulate our feature coding model, we start from the reconstruction error. Specifically, each person is expected to be well represented by the learned dictionary, and according to the definitions of SDL and ADL, the reconstruction model can be typically formulated as follows [28]:

$$\min_{\mathbf{D},\mathbf{P}} ||\mathbf{X} - \mathbf{D}\mathbf{P}\mathbf{X}||_F^2 \;\text{ s.t. } \forall\, i,\; ||\mathbf{d}_i||_2 \leq 1. \tag{4}$$

We now focus on our discriminative term. In the person re-identification task, it usually does not provide the strong class labels, but the weak pairwise labels, i.e., the same person pairs and the different person pairs. To utilize such the discriminative information, we assume that *each transformed data sample* $\mathbf{P}\mathbf{x}_i$ *has maximum dependence to the ones from the same person pairs and minimum dependence to the ones from different persons*. Thus, we can employ HSIC to address it. Specifically, in the transformed data space, we adopt the linear inner product $\mathbf{K} = (\mathbf{P}\mathbf{X})^T\mathbf{P}\mathbf{X}$ as its kernel. In the weak pairwise label space, we define a new kernel matrix $\mathbf{W}$. Obviously, such a new kernel matrix $\mathbf{W}$ should satisfy that $w_{ij} \geq 0$ when the samples $\mathbf{x}_i$ and $\mathbf{x}_j$ are from the same person, and $w_{ij} \leq 0$ otherwise. Besides this, two additional properties are also beneficial. One is that the kernel matrix $\mathbf{W}$ should be symmetric (i.e., $w_{ij} = w_{ji}$), which means that the dependence/similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$ is undirected. The other one is that the matrix $\mathbf{W}$ should satisfy $\forall i,\ \sum_j^n w_{ij} = 0$, which is to balance the contributions of same person pairs and different person pairs because the number of different person pairs is in general much larger than the same ones. Based on these properties, we simply define the matrix $\mathbf{W}$ as follows:

$$w_{ij} = \begin{cases} 1/n_k, & (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S} \\ -1/(n - n_k), & (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D} \end{cases} \tag{5}$$

where $\mathcal{S}$ means that $\mathbf{x}_i$ and $\mathbf{x}_j$ are from the same person, $\mathcal{D}$ otherwise. n_k denotes the number of samples from the kth person. Obviously, this definition satisfies the above three properties. Thus, we can exploit the discriminative term as

$$\max_{\mathbf{P}} \mathrm{tr}((\mathbf{P}\mathbf{X})^T\mathbf{P}\mathbf{X}\mathbf{H}\mathbf{W}\mathbf{H}) = \min_{\mathbf{P}} \mathrm{tr}(\mathbf{P}\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P}^T) \tag{6}$$

where the data $\mathbf{X}$ is centered (i.e., $\mathbf{X} = \mathbf{X}\mathbf{H}$) and $\mathbf{L} = -\mathbf{W}$. Putting every concern together, say (4) and (6), the proposed dependence-aware feature coding (DAFC) model turns out to be like

$$\begin{aligned} \min_{\mathbf{D},\mathbf{P}} \;& ||\mathbf{X} - \mathbf{D}\mathbf{P}\mathbf{X}||_F^2 + \lambda \mathrm{tr}(\mathbf{P}\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P}^T) \\ \text{s.t. } & \forall\, i,\; ||\mathbf{d}_i||_2 \leq 1,\; ||\mathbf{p}^i||_2 \leq 1 \end{aligned} \tag{7}$$

where λ is the tradeoff parameter. The constraints $\{\forall\, i,\ ||\mathbf{d}_i||_2 \leq 1, ||\mathbf{p}^i||_2 \leq 1\}$ are to avoid the scale issue.

IV. Optimization

For the proposed model (7), it is generally not a jointly convex optimization problem for $\{\mathbf{D}, \mathbf{P}\}$, but is convex with respect to each variable.[1] Therefore, we adopt the alternative convex search (ACS) [35] to address it. To make the objective function easy to solve, we introduce an auxiliary variable $\mathbf{Z}$ to make all the subproblems separable. In the sequel, the objective (7) becomes the following optimization problem:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{D},\mathbf{P}} \quad & ||\mathbf{X} - \mathbf{D}\mathbf{Z}||_F^2 + \tau||\mathbf{P}\mathbf{X} - \mathbf{Z}||_F^2 + \lambda \mathrm{tr}(\mathbf{P}\mathbf{X}\mathbf{L}(\mathbf{P}\mathbf{X})^T) \\ \text{s.t.} \quad & \forall\, i,\; ||\mathbf{d}_i||_2 \leq 1,\; ||\mathbf{p}^i||_2 \leq 1 \end{aligned} \tag{8}$$

[1] We will analyze this fact in the next section.

Algorithm 1: DAFC.

Input: Centered training samples $\mathbf{X} \in \mathbb{R}^{d\times n}$, $\mathbf{D}_{init}$, $\mathbf{P}_{init}$, kernel matrix $\mathbf{W}$, dictionary size m, parameter λ.
Output: Discriminative analysis dictionary $\mathbf{P}$.
while *unreached the terminal condition* **do**
  Update $\mathbf{Z}$ via (9);
  Update $\mathbf{D}$ via (11);
  Update $\mathbf{P}$ via (12);
end

where τ is a positive scalar constant. Hence, there are three variables, including $\mathbf{Z}$, $\mathbf{D}$, and $\mathbf{P}$, to solve.

Z-subproblem: Taking derivative of the objective with respect to $\mathbf{Z}$ and setting it to zero reads

$$\begin{aligned}\mathbf{Z} &= \text{argmin}_{\mathbf{Z}} \ ||\mathbf{X}-\mathbf{DZ}||_F^2 + \tau||\mathbf{PX}-\mathbf{Z}||_F^2 \\ &= (\mathbf{D}^T\mathbf{D}+\tau\mathbf{I})^{-1}(\tau\mathbf{PX}+\mathbf{D}^T\mathbf{X}).\end{aligned} \tag{9}$$

D-subproblem: By discarding the unrelated terms to $\mathbf{D}$

$$\min_{\mathbf{D}} \ ||\mathbf{X}-\mathbf{DZ}||_F^2 \ \text{ s.t. } \forall i, \ \ ||\mathbf{d}_i||_2 \ \leq 1. \tag{10}$$

By Alternating Direction Method of Multipliers (ADMM) algorithm [28], [36], the optimal solution is

$$\begin{cases}\mathbf{D}^{(r+1)} = \text{argmin}_{\mathbf{D}} ||\mathbf{X}-\mathbf{DZ}||_F^2 + \rho||\mathbf{D}-\mathbf{S}^{(r)}+\mathbf{T}^{(r)}||_F^2 \\ \qquad\quad = (\mathbf{XZ}^T + \rho(\mathbf{S}^{(r)}-\mathbf{T}^{(r)}))(\mathbf{ZZ}^T+\rho\mathbf{I})^{-1} \\ \mathbf{S}^{(r+1)} = \text{argmin}_{\mathbf{S}} \rho||\mathbf{D}^{(r+1)}-\mathbf{S}+\mathbf{T}^{(r)}||_F^2, \text{s.t. } ||\mathbf{s}_i||_2^2 \leq 1 \\ \mathbf{T}^{(r+1)} = \mathbf{T}^{(r)} + \mathbf{D}^{(r+1)} - \mathbf{S}^{(r+1)}\end{cases} \tag{11}$$

where ρ is a penalty scalar and is updated if appropriate.

P-*subproblem:* Fixing the other variables gives

$$\begin{aligned}&\min_{\mathbf{P}} \ \ \tau||\mathbf{PX}-\mathbf{Z}||_F^2 + \lambda\text{tr}(\mathbf{PXLX}^T\mathbf{P}^T) \\ &\quad \text{s.t. } \forall i, \ \ ||\mathbf{p}^i||_2 \leq 1.\end{aligned} \tag{12}$$

Similarly, this optimization can also be solved by ADMM.

The entire algorithm of DAFC is summarized in Algorithm 1, which terminates when the relative change of objective value between two neighboring iterations is sufficiently small ($|\frac{f(t+1)-f(t)}{f(t)}| \leq 10^{-3}$) or the maximal iterative number ($T=100$) is reached. For the initializations, similar to [32], we use *k*-means to initialize the synthesis dictionary $\mathbf{D}_{\text{init}}$. The analysis dictionary $\mathbf{P}_{\text{init}}$ is directly assigned as $\mathbf{D}_{\text{init}}^T$.

V. Theoretical Analysis

According to the ACS algorithm [35], each subproblem of DAFC need to be convex. It is easy to verify that the subproblems of $\mathbf{Z}$ and $\mathbf{D}$ are convex. For the subproblem of $\mathbf{P}$, i.e., (12), obviously, its constraints $\{\forall i, \ ||\mathbf{p}^i||_2 \leq 1\}$ are convex and its objective function is

$$f(\mathbf{P}) = \tau||\mathbf{PX}-\mathbf{Z}||_F^2 \ + \lambda\text{tr}(\mathbf{PXLX}^T\mathbf{P}^T). \tag{13}$$

The term $\text{tr}(\mathbf{PXLX}^T\mathbf{P}^T)$ is generally nonconvex and unstable due to the nonpositive similarity values involved. This leads to another question: Is the holistic function $f(\mathbf{P})$ convex? Before answering it, we would like to prove a theorem.

Lemma 1 (Gerschgorin theorem [37]): Let $\mathbf{A}=[a_{ij}]$ be an arbitrary $n\times n$ complex matrix, and let $R_i = \sum_{j=1;j\neq i}^n |a_{ij}|, \ \ 1 \leq i \leq n$, where $R_i := 0$ if $n=1$. If λ is an eigenvalue of $\mathbf{A}$, then there is a positive integer r, with $1 \leq r \leq n$, such that

$$|\lambda - a_{rr}| \leq R_r \tag{14}$$

Hence, all eigenvalues λ of $\mathbf{A}$ lie in the union of the disks.

We refer readers to the work [37] for the detailed proof of Lemma 1. With Lemma 1, we can prove the following theorem.

Theorem 1: For a matrix $\mathbf{B}=\mathbf{A}+\alpha\mathbf{I} \in \mathbb{R}^{n\times n}$, where $\mathbf{A}=[a_{ij}]$ is an arbitrary $n\times n$ complex matrix and α is a nonnegative value. $\mathbf{B}$ is semipositive definite when the parameter α satisfies the following constraint:

$$\alpha \geq \max_{1\leq i\leq n}\left(\sum_{j=1;j\neq i}^{n}|a_{ij}| - a_{ii}\right). \tag{15}$$

Proof: To make the matrix $\mathbf{B}$ to be semipositive definite, according to the work [38], it means that the minimal eigenvalue of $\mathbf{B}$ needs to be nonnegative. Fortunately, based on the Lemma 1, we know that all the eigenvalues η of $\mathbf{B}$ lie in $|\eta - a_{ii} - \alpha| \leq \sum_{j=1;j\neq i}^n |a_{ij}|, \ \ 1\leq i \leq n$. To make the minimal eigenvalue $\eta_{\min}$ to always be nonnegative, after some transformations, it is easy to verify that the value of α has to satisfy the constraint (15). ∎

To answer the above question, we know that the convexity of $f(\mathbf{P})$ depends on whether its Hessian matrix $\nabla^2 f(\mathbf{P})$ is semipositive definite or not [38]. Fortunately, the Hessian matrix $\nabla^2 f(\mathbf{P})$ can be easily computed as follows:

$$\nabla^2 f(\mathbf{P}) = \lambda\mathbf{XLX}^T + \tau\mathbf{XX}^T. \tag{16}$$

Let $\mathbf{C}=\mathbf{L}+\tau\mathbf{I}/\lambda$. To guarantee the Hessian matrix $\nabla^2 f(\mathbf{P}) = \lambda\mathbf{XCX}^T$ to be semipositive definite, $\mathbf{C}$ should be semipositive definite. Thus, we can obtain $\tau/\lambda \geq \max_{1\leq i\leq n}(\sum_{j=1;j\neq i}^n |l_{ij}| - l_{ii})$. More concretely, according to the replacement $\mathbf{L}=-\mathbf{W}$ and the definition of $\mathbf{W}$ [i.e., (5)], we know that the lower bound of τ/λ is 2. Thus, we set $\tau = 2\lambda$ in all the experiments.

In this way, we know that each subproblem of our DAFC is convex. By fixing $\mathbf{Z}$, the variables $\mathbf{D}$ and $\mathbf{P}$ are separable, and they can be termed as a single variable. Thus, the optimization problem in (8) is a biconvex problem of $\min_{\mathbf{Z}}\{f(\mathbf{Z},(\mathbf{D}_t,\mathbf{P}_t))\}$ and $\min_{(\mathbf{D},\mathbf{P})}\{f(\mathbf{Z}_t,(\mathbf{D},\mathbf{P}))\}$. In the training, we alternatively solve the two convex optimization problems, and the whole function $f(\mathbf{Z},(\mathbf{D},\mathbf{P}))$ generally has a lower bound. Therefore, according to the ACS algorithm [35], we know that the proposed DAFC algorithm is guaranteed to converge monotonically in terms of objective value.

VI. Experiments

We strictly follow all the experimental settings as the work [32], including the adopted datasets, the data processing method, and the evaluation criteria.

Dataset description: In this section, we apply the proposed DAFC algorithm on the person re-identification task. Two publicly available VIPeR and PRID450S datasets are adopted in this letter. VIPeR dataset is composed of 632 persons and each person has two images captured in outdoor environments. It mainly suffers from arbitrary viewpoints and illumination

TABLE I
TOP-RANKED MATCHING RATES ON VIPeR DATASET, COMPARED WITH DIFFERENT TYPICAL CODING METHODS. BEST IN BOLD

	Rank	1	5	10	20
Unsupervised	SAC [21]	0.393	0.695	0.811	0.901
	LLC(5) [22]	0.128	0.325	0.456	0.609
	LLC(120) [22]	0.393	0.696	0.813	0.902
	SC(5) [23]	0.116	0.313	0.453	0.615
	SC(120) [23]	0.395	0.700	0.816	0.904
Supervised	CPDL [33]	0.360	0.642	0.755	0.843
	KSDL [25]	0.392	0.684	0.807	0.898
	DPL [28]	0.394	0.698	0.812	0.902
	DADL [31]	0.404	0.703	0.825	0.901
	MEDVL [32]	0.411	0.717	0.832	**0.917**
Our	DAFC	**0.449**	**0.743**	**0.841**	0.914

TABLE II
TOP-RANKED MATCHING RATES ON PRID450S DATASET, COMPARED WITH DIFFERENT TYPICAL CODING METHODS. BEST IN BOLD

	Rank	1	5	10	20
Unsupervised	SAC [21]	0.434	0.704	0.805	0.890
	LLC(5) [22]	0.094	0.278	0.406	0.569
	LLC(120) [22]	0.433	0.706	0.805	0.891
	SC(5) [23]	0.085	0.265	0.397	0.560
	SC(120) [23]	0.440	0.713	0.814	0.898
Supervised	CPDL [33]	0.380	0.670	0.765	0.869
	KSDL [25]	0.421	0.698	0.798	0.882
	DPL [28]	0.429	0.704	0.802	0.889
	DADL [31]	0.443	0.721	0.816	0.907
	MEDVL [32]	0.459	0.730	0.829	0.911
Our	DAFC	**0.465**	**0.744**	**0.847**	**0.915**

variations between two disjoint cameras. PRID450S dataset is another challenging dataset that is captured with different view changes, background interference, and occlusion variations and consists of 450 person pairs.

Raw features: We employ the image-level features provided by the work [32] as the inputs. Moreover, as suggested by the work [32], the dimensions of the image-level features of both two datasets are reduced to 70 by Principal Components Analysis (PCA).

Settings: Following the standard protocol in [32], with the learned analysis dictionary $\mathbf{P}$, we can directly obtain the coding features as $\mathbf{PX}$. Then, the metric learning method KISSME [13] is employed to guide the final person pairs matching. For both datasets, half person pairs are randomly selected as the training set and the remaining as testing. For the evaluation criterion, the average of Rank-k recognition rates over ten independent runs are reported.

A. Comparison With Different Coding Methods

To validate whether the high-level features learned by our coding methods are more discriminative than other alternatives or not, we employ three unsupervised coding methods including SAC [21], LLC [22], and SC [23] to encode the input image-level features. For these methods, the k-means technique is used to construct the dictionary. Moreover, four supervised dictionary learning methods including Kernelized Supervised Dictionary Learning (KSDL) [25], Dictionary Pair Learning (DPL) [28], Discriminative Analysis Dictionary Learning (DADL) [31], and Metric Embedding Discriminative Vocabulary Learning (MEDVL) [32] and one supervised Cross-view Pair Dictionary Learning (CPDL) method [33] are[2] are also compared to show the advantages of our method. To all the compared methods, their source codes can be downloaded from the github or from authors' webpages, and the parameters are tuned according to their suggestions. From Tables I and II, we can observe that all unsupervised coding methods including SAC, LLC(120),[3] and SC(120) perform relatively promising. Among them, the SC(120) seems to be the best. This is consistent with the intuition that the saliency of person images is important in the person re-identification task. Additionally, it notes that on both datasets, LLC(5) and SC(5) perform poorly; this means that when encoding the image-level features into the high-level semantic ones, the locality constraint may harm the performance of person re-identification. Compared the supervised method MEDVL with these unsupervised competitors, it can be observed that MEDVL generally performs better, due to the involved weak pairwise labels. For the compared method DADL, its performance is promising but lower than the alternative MEDVL. For the method CPDL, it only uses the positive person pairs as the discriminative regularization and its performance is lower than MEDVL. For the competitor DPL, as it is designed for the multiclass classification task, its performance in the person re-identification task is not promising, and is comparable with the unsupervised methods, but is lower than the method MEDVL. For the compared method KSDL, it only uses the pairs from the same person and ignores the different ones. Its performance is also limited in the person re-identification task. For the proposed method DAFC, it employs the HSIC as the discriminative term. From the experimental results in the tables, we can clearly see that our method outperforms all the competitors in most of cases. We have achieved about 5.4% Rank 1 improvement on VIPeR and 2.5% Rank 1 improvement on PRID450S over the most promising unsupervised competitor SC(120). In addition, compared with the supervised coding method MEDVL, we can achieve about 3.8% Rank 1 improvement on VIPeR and 0.6% Rank 1 improvement on PRID450S. Therefore, we can conclude that the employed HSIC is more discriminative than the previous alternatives and is more suitable for the person re-identification task. In all the experiments, the parameters are chosen by ten-fold cross validation. The best parameter λ on VIPeR and PRID450S datasets is 0.6 and 0.5, respectively. The running times of the proposed DAFC on VIPeR and PRID450S datasets are 6.5 and 3.5 s, respectively.

VII. CONCLUSION

In this letter, we have proposed a novel dDFAC framework, which employs the HSIC as a regularization to improve the dictionary discriminability, and is applicable to the person re-identification task. Moreover, theoretical discussion and analysis on the convexity of the proposed constraint, as well as the convergence of DAFC algorithm, are provided. Experimental results on two benchmark datasets VIPeR and PRID450S have shown the advantages of our method over the state-of-the-art alternatives.

[2]We only use the image-level part of CPDL [33] for fair comparison.
[3](5) and (120) are the corresponding dictionary sizes.

REFERENCES

[1] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Comput. Surveys*, vol. 46, no. 2, 2013, Art. no. 29.

[2] B. Song, A. T. Kamal, and C. Soto, "Tracking and activity recognition through consensus in distributed camera networks," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2564–2579, Oct. 2010.

[3] M. Hirzer, P. M. Roth, and M. Kostinger, "Relaxed pairwise learned metric for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 780–793.

[4] C. Tian, M. Zeng, and Z. Wu, "Person re-identification based on spatiogram descriptor and collaborative representation," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1595–1599, Oct. 2015.

[5] Y. Xie, H. Yu, and X. Gong, "Learning visual-spatial saliency for multiple-shot person re-identification," *IEEE Signal Process. Letters*, vol. 22, no. 11, pp. 1854–1858, Nov. 2015.

[6] Y. Chen, W. Zheng, and J. Lai, "Mirror representation for modeling view-specific transform in person re-identification," in *Proc. 24th Int. Conf. Artif. Intell.*, 2015, pp. 3402–3408.

[7] X. Jing, X. Zhu, F. Wu, and X. You, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 695–704.

[8] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1116–1124.

[9] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 1268–1277.

[10] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 3625–3661.

[11] W. Li, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017.

[12] K. Liu, Z. Zhao, and A. Cai, "Datum-adaptive local metric learning for person re-identification," *IEEE Signal Process. Letters*, vol. 22, no. 9, pp. 1457–1461, Sep. 2015.

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

[14] M. Farenzena, L. Bazzani, and A. Perina, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 2360–2367.

[15] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1622–1634, Jul. 2013.

[16] Y. Yang, J. Yang, and J. Yan, "Salient color names for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 536–551.

[17] R. Zhao, W. Oyang, and X. Wang, "Person re-identification by saliency learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 356–370, Feb. 2017.

[18] S. Iodice and A. Petrosino, "Salient feature based graph matching for person re-identification," *Pattern Recog.*, vol. 48, no. 4, pp. 1074–1085, 2015.

[19] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2014, pp. 144–151.

[20] Y. Yang, S. Liao, and Z. Lei, "Color models and weighted covariance estimation for person re-identification," in *Proc. Int. Conf. Pattern Recog.*, 2014, pp. 1874–1879.

[21] G. Van, J. M. Geusebroek, and C. J. Veenman, "Kernel codebooks for scene categorization," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 696–709.

[22] J. Wang, J. Yang, and K. Yu, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010.

[23] Y. Huang, K. Huang, and Y. Yu, "Salient coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 1753–1760.

[24] H. Guo, Z. Jiang, and L. S. Davis, "Discriminative dictionary learning with pairwise constraints," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 328–342.

[25] M. J. Gangeh, A. Ghodsi, and M. S. Kamel, "Kernelized supervised dictionary learning," *IEEE Trans. Signal Process.*, vol. 61, no, 19, pp. 4753–4767, Oct. 2013.

[26] E. M. Eksioglu and O. Bayir, "K-SVD meets transform learning: Transform K-SVD," *IEEE Signal Process. Lett.*, vol. 21, no. 3, pp. 347–351, Mar. 2014.

[27] V. Abolghasemi, M. Chen, and A. Alameer, "Incoherent dictionary pair learning: Application to a novel open-source database of chinese numbers," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 472–276, Apr. 2018.

[28] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proc. Advances Neural Inform. Process. Syst.*, 2014, pp. 793–801.

[29] J. Wang, Y. Guo, and J. Guo, "Class-aware analysis dictionary learning for pattern classification," *IEEE Signal Process. Letters*, vol. 24, no. 12, pp. 1822–1826, Dec. 2017.

[30] M. Yang, H. Chang, and W. Luo, "Discriminative analysis-synthesis dictionary learning for image classification," *Neurocomputing*, vol. 219, pp. 404–411, 2017.

[31] J. Guo, Y. Guo, X. Kong, M. Zhang, and R. He, "Discriminative analysis dictionary learning," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 1617–1623.

[32] Y. Yang, L. Zhen, S. Zhang, H. Shi, and S. Li, "Metric embedded discriminative vocabulary learning for high-level person representation," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 3648–3654.

[33] S. Li, M. Shao, and Y. Fu, "Cross-view projective dictionary learning for person re-identification," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 2155–2161.

[34] A. Gretton, O. Bousquet, and A. Smola, "Measuring statistical dependence with Hilbert-Schmidt norms," in *Proc. Int. Conf. Algorithmic Learning Theory*, 2005, pp. 63–77.

[35] W. Richard and A. Hurter, "Minimization of a non-separable objective function subject to disjoint constraints," *Oper. Res.*, vol. 24, no. 4, pp. 643–957, 1976.

[36] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 612–620.

[37] R. Varga, *Matrix Iterative Analysis*. Berlin, Germany: Springer-Verlag, 2009.

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