COUPLED ANALYSIS-SYNTHESIS DICTIONARY LEARNING FOR PERSON RE-IDENTIFICATION

Lingchuan Sun[1], Yun Zhou [2], Zhuqing Jiang[1], Aidong Men[1]

[1]Beijing University of Posts and Telecommunications, Beijing, China
[2]Academy of Broadcasting Science, Beijing, China
{sunlingchuan, jiangzhuqing, menad}@bupt.edu.cn, zhouyun@abs.ac.cn

ABSTRACT

In this paper, we propose a novel coupled dictionary learning method, namely coupled analysis-synthesis dictionary learning, to improve the performance of person re-identification in the non-overlapping fields of different camera views. Most of the existing coupled dictionary learning methods train a coupled synthesis dictionary directly on the original feature spaces, which limits the representation ability of the dictionary. To handle the diversities of different original spaces, We first employ local Fisher discriminant analysis (LFDA) to learn a common feature space for close relationship of the same people in different views. In order to enhance the representation power of the coupled synthesis dictionary, we then learn a coupled analysis dictionary by transforming the common feature space into the coupled feature space. Experimental results on two publicly available VIPeR and CUHK01 datasets have validated the effectiveness of the proposed method.

***Index Terms*—** Person re-identification, LFDA, coupled dictionary learning, analysis-synthesis dictionary

1. INTRODUCTION

Person re-identification is the task of matching specific person across non-overlapping camera views. It has received increasing attentions in recent years [1]. Despite many approaches, person re-identification is still a challenge due to occlusion, viewpoint, body pose and illumination changes in different camera views.

Most of researches for addressing this challenge mainly focus on two aspects: feature representation and distance metric learning. Feature representation designs discriminative and robust descriptors to distinguish different persons [2, 3, 4]. Distance metric learning aims to learn distance measure that makes the same person more similar and the different people more different [4, 5, 6, 7]. Recently, deep learning and manifold learning have shown great potential in person re-identification [8, 9, 10]. However, the various viewing conditions of the person images limit the representational and discriminative power of these methods.

In addition, dictionary learning has shown the strong representational power and been used to the cross-view problems [11, 12, 13]. Specially, coupled dictionary learning (CDL) has also proved effective for person re-identification [14, 15, 16]. Liu et al. [14] jointly learned a coupled dictionary with labeled and unlabeled images. Jing et al. [15] jointly learned a semi-coupled dictionary and a mapping function from HR gallery images and LR probe images. For person re-identification, the person features from one view form a original feature space. Most of the existing CDL methods train a coupled dictionary directly in these original feature spaces. However, different viewpoints and correlation structures between original feature spaces restrict the representation capability of the dictionary. Moreover, traditional CDL adopts l_0 or l_1-norm to regularize the representation coefficients in training process, which results in the heavy computational burden and makes the training and testing inefficient [17]. Recently, analysis-synthesis dictionary learning has attracted much attention [18]. Gu et al. [17] proposed a projective dictionary pair learning (PDPL), which learns a synthesis dictionary for signal representation and an analysis dictionary for discrimination coding simultaneously.

In this paper, we first use local Fisher discriminant analysis (LFDA) [19] to learn a common feature space for capturing the intrinsic structures of cross-view data. Then we learn discriminative and robust representations by extending the PDPL method. The major contributions in this paper include: (1) We propose a coupled analysis-synthesis dictionary learning model, which is more efficient than traditional coupled dictionary learning. This model trains a coupled synthesis dictionary for robust feature representation and a coupled analysis dictionary for discriminative coding. (2) To improve the representation ability of the coupled synthesis dictionary, we construct an associate function with coupled analysis dictionary which is capable of transforming the common feature space into the coupled feature space. An efficient optimization algorithm is derived to solve the model.

The rest of the paper is organized as follows: Section 2

This work is supported by the Project of the National Natural Science Foundation of China No.61671077 and No.61671264.

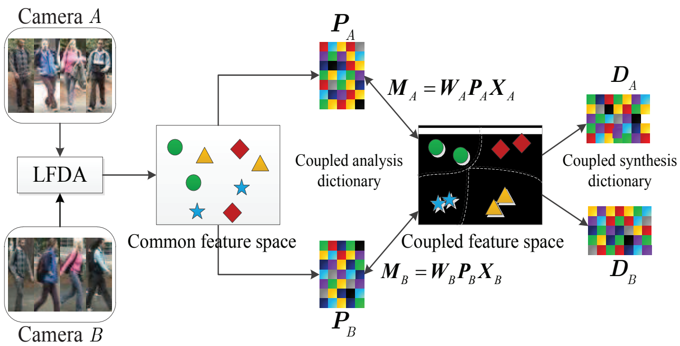

Fig. 1: Illustration of our proposed training framework for person re-identification. Note $\{\mathbf{P}_A, \mathbf{P}_B\}$ are used to code representation coefficients, $\{\mathbf{W}_A, \mathbf{W}_B\}$ are used to transform coding coefficients into coupled feature space, $\{\mathbf{D}_A, \mathbf{D}_B\}$ are used to reconstruct $\{\mathbf{X}_A, \mathbf{X}_B\}$.

describes our method in detail, Section 3 presents experimental results. Finally, Section 4 concludes this paper.

2. PROPOSED METHOD

In this section, we first present the scheme of our algorithm, including learning common feature space by LFDA and coupled analysis-synthesis dictionary learning. Then optimization details for our dictionary learning model are presented. Finally, matching method is introduced for person re-identification.

2.1. Learning Common Feature Space by LFDA

LFDA can not only project data into a common feature space to minimize the within-class scatter and maximize the between-class scatter, but also preserve the local structure of data by using locality preserving projection. Hence, it is capable of capturing the close relationship of the same person in different views.

Given n training person image pairs from two non-overlapping camera A and B, image features from A and B form two different original feature spaces. LFDA is utilized to learn a transformation matrix $\mathbf{T}$ on training set that tansforms all original features into a common feature space. The details are shown in [19].

2.2. Coupled Analysis-Synthesis Dictionary Learning

Let $x \in R^m$ stand for a m-dimentional feature vector from the learned common feature space. $\mathbf{X}_A = \left[x_A^1, ..., x_A^i, ...x_A^n\right] \in \mathbf{R}^{m \times n}$, $i = 1, 2, ...n$ and $\mathbf{X}_B = \left[x_B^1, ..., x_B^i, ...x_B^n\right] \in \mathbf{R}^{m \times n}$, $i = 1, 2, ...n$ respectively represent training data matrices captured by camera A and B, where x_A^i and x_B^i mean the same person under the different camera views. We denote $\mathbf{P}_A \in \mathbf{R}^{d_1 \times m}$(or $\mathbf{P}_B \in \mathbf{R}^{d_2 \times m}$) as analysis dictionary and $\mathbf{D}_A \in \mathbf{R}^{m \times d_1}$(or $\mathbf{D}_B \in \mathbf{R}^{m \times d_2}$) as synthesis dictionary. The coupled analysis-synthesis dictionary learning on the common feature space can be formulated under the following framework:

$$\begin{aligned}\min_{\mathbf{D}_A, \mathbf{D}_B, \mathbf{P}_A, \mathbf{P}_B} & \|\mathbf{X}_A - \mathbf{D}_A \mathbf{P}_A \mathbf{X}_A\|_F^2 + \|\mathbf{X}_B - \mathbf{D}_B \mathbf{P}_B \mathbf{X}_B\|_F^2 \\ & + \Psi\left(\mathbf{P}_A, \mathbf{P}_B\right) \\ & s.t. \quad \|d_{A,i}\|_2^2 \leq 1, \|d_{B,i}\|_2^2 \leq 1, \forall i\end{aligned} \tag{1}$$

Where $d_{A,i}$ and $d_{B,i}$ denote i-th atom of $\mathbf{D}_A$ and $\mathbf{D}_B$, respectively. $\Psi\left(\mathbf{P}_A, \mathbf{P}_B\right)$ stands for the associate function which describes the relationship between two views. In the first term of Eq.(1), $\mathbf{P}_A$ and $\mathbf{D}_A$ form an analysis-synthesis dictionary pair in the view A. By joint learning $\mathbf{P}_A$ and $\mathbf{D}_A$, the coding coefficient matrix corresponding to the feature set $\mathbf{X}_A$ can be analytically obtained as $\mathbf{Z}_A = \mathbf{P}_A \mathbf{X}_A$, and $\mathbf{X}_A$ can be reconstructed as $\mathbf{X}_A \approx \mathbf{D}_A \mathbf{Z}_A$ using synthesis dictionary $\mathbf{D}_A$ simultaneously. The same as above, we can analytically obtain the coding coefficient matrix as $\mathbf{Z}_B = \mathbf{P}_B \mathbf{X}_B$ and reconstruct $\mathbf{X}_B$ with $\mathbf{D}_B$ in the view B. In our work, we aim to define the association function which is able to better describe the intrinsic structures of cross-view data.

Since there are viewpoint invariant structures of the same object in different views, CDL makes the reasonable assumption that there exists a latent coupled feature space where the coding coefficients of the same object should be strictly equal. However, this assumption is too strong to handle various changes of image structures from different views. In this paper, we relax this assumption, following the scheme in [20], and employ a pair of analysis dictionaries to leran the coupled feature space.

Let $\mathbf{M}_A = \mathbf{W}_A \mathbf{Z}_A$ be the projected data from analytical coding coefficient matrix $\mathbf{Z}_A$, where $\mathbf{W}_A \in \mathbf{R}^{d_c \times d_1}$ is the mapping transform. The same remarks are applied to $\mathbf{M}_B$ and $\mathbf{W}_B$. By learning $\mathbf{W}_A$ and $\mathbf{W}_B$, $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are transformed into a coupled feature space. Therefore, we consider the following minimization problem:

$$\begin{aligned}\min \|\mathbf{M}_A - \mathbf{M}_B\|_F^2 &= \min \|\mathbf{W}_A \mathbf{Z}_A - \mathbf{W}_B \mathbf{Z}_B\|_F^2 \\ &= \min \|\mathbf{W}_A \mathbf{P}_A \mathbf{X}_A - \mathbf{W}_B \mathbf{P}_B \mathbf{X}_B\|_F^2\end{aligned} \tag{2}$$

It is worth to point out that $\mathbf{W}$ and $\mathbf{P}$ are different in Eq.(2). $\mathbf{W}$ is used to map coding coefficient matrix $\mathbf{Z}$ into a coupled space, and the analysis dictionary $\mathbf{P}$ is used to analytically code $\mathbf{X}$. To avoid the trivial solution and be more precise, for a same person $m = \mathbf{W}_A \mathbf{P}_A x_A^i$ (or $m = \mathbf{W}_B \mathbf{P}_B x_B^i$) in the coupled feature space, we can derive $\mathbf{P}_B x_B^i = \mathbf{W}_B^{-1} m$ (or $\mathbf{P}_A x_A^i = \mathbf{W}_A^{-1} m$) so that this person in the other camera view can be reconstructed by $\mathbf{D}_B \mathbf{P}_B x_B^i$ (or $\mathbf{D}_A \mathbf{P}_A x_A^i$). Finally, the associate function $\Psi\left(\mathbf{P}_A, \mathbf{P}_B\right)$ can be formulated as :

$$\begin{aligned}\Psi\left(\mathbf{P}_A, \mathbf{P}_B\right) = & \left\|\mathbf{P}_A \mathbf{X}_A - \mathbf{W}_A^{-1} \mathbf{M}_B\right\|_F^2 \\ & + \left\|\mathbf{P}_B \mathbf{X}_B - \mathbf{W}_B^{-1} \mathbf{M}_A\right\|_F^2\end{aligned} \tag{3}$$

We substitute Eq.(3) into Eq.(1), the objective function of coupled analysis-synthesis dictionary learning is formulated below:

$$\begin{gathered}\min_{\substack{\mathbf{D}_A,\mathbf{D}_B,\mathbf{P}_A,\mathbf{P}_B,\\ \mathbf{W}_A,\mathbf{W}_B}} \|\mathbf{X}_A-\mathbf{D}_A\mathbf{P}_A\mathbf{X}_A\|_F^2+\|\mathbf{X}_B-\mathbf{D}_B\mathbf{P}_B\mathbf{X}_B\|_F^2\\ +\lambda_1\left(\left\|\mathbf{P}_A\mathbf{X}_A-\mathbf{W}_A^{-1}\mathbf{M}_B\right\|_F^2+\left\|\mathbf{P}_B\mathbf{X}_B-\mathbf{W}_B^{-1}\mathbf{M}_A\right\|_F^2\right)\\ +\lambda_2\left(\left\|\mathbf{W}_A^{-1}\right\|_F^2+\left\|\mathbf{W}_B^{-1}\right\|_F^2\right)\\ s.t.\quad \|d_{A,i}\|_2^2\le 1, \|d_{B,i}\|_2^2\le 1, \forall i\end{gathered}\tag{4}$$

Where λ_1 and λ_2 are scalar parameters, the additionally regularized constraints on $\mathbf{W}_A^{-1}$ and $\mathbf{W}_B^{-1}$ are imposed to avoid over-fitting. In Eq.(4), once mapping $\mathbf{W}_A^{-1}$ and $\mathbf{W}_B^{-1}$ are obtained, the coupled analysis-synthesis dictionary can be updated to better analysis coding coefficients and reconstruct the person images. Fig.1 shows the training framework of our method.

2.3. Optimization

To solve the optimization problem of Eq.(4), we introduce auxiliary variables $\mathbf{Z}_A$ and $\mathbf{Z}_B$ and relax the objective function. Hence the function can be rewritten as follows:

$$\begin{gathered}\min_{\substack{\mathbf{D}_A,\mathbf{D}_B,\mathbf{P}_A,\mathbf{P}_B,\\ \mathbf{W}_A,\mathbf{W}_B,\mathbf{Z}_A,\mathbf{Z}_B}} \|\mathbf{X}_A-\mathbf{D}_A\mathbf{Z}_A\|_F^2+\|\mathbf{X}_B-\mathbf{D}_B\mathbf{Z}_B\|_F^2\\ +\tau\left(\|\mathbf{P}_A\mathbf{X}_A-\mathbf{Z}_A\|_F^2+\|\mathbf{P}_B\mathbf{X}_B-\mathbf{Z}_B\|_F^2\right)\\ +\lambda_1\left(\left\|\mathbf{Z}_A-\mathbf{W}_A^{-1}\mathbf{M}_B\right\|_F^2+\left\|\mathbf{Z}_B-\mathbf{W}_B^{-1}\mathbf{M}_A\right\|_F^2\right)\\ +\lambda_2\left(\left\|\mathbf{W}_A^{-1}\right\|_F^2+\left\|\mathbf{W}_B^{-1}\right\|_F^2\right)\\ s.t.\quad \|d_{A,i}\|_2^2\le 1, \|d_{B,i}\|_2^2\le 1, \forall i\end{gathered}\tag{5}$$

Where τ is regularization parameter. Although there are many variables in Eq.(5), it can be solved by alternatively optimizing these variables until convergence. This iterative update process is described in the following part.

Updating $\mathbf{Z}_A$ and $\mathbf{Z}_B$ We first fix other variables, and update variables $\mathbf{Z}_A$ and $\mathbf{Z}_B$. For $\mathbf{Z}_A$, the objective function can be simplified as:

$$\begin{gathered}\min_{\mathbf{Z}_A}\|\mathbf{X}_A-\mathbf{D}_A\mathbf{Z}_A\|_F^2+\tau\|\mathbf{P}_A\mathbf{X}_A-\mathbf{Z}_A\|_F^2\\ +\lambda_1\left\|\mathbf{Z}_A-\mathbf{W}_A^{-1}\mathbf{M}_B\right\|_F^2\end{gathered}\tag{6}$$

The closed-form solution of $\mathbf{Z}_A$ can be derived as:

$$\begin{aligned}\mathbf{Z}_A=&\left(\mathbf{D}_A^T\mathbf{D}_A+(\lambda_1+\tau)I\right)^{-1}\\ &\left(\mathbf{D}_A^T\mathbf{X}_A+\lambda_1\mathbf{W}_A^{-1}\mathbf{M}_B+\tau\mathbf{P}_A\mathbf{X}_A\right)\end{aligned}\tag{7}$$

Where I is an identity matrix. we derive $\mathbf{Z}_B$ in a similar way.

Updating $\mathbf{W}_A$ and $\mathbf{W}_B$ We only consider the terms associated with $\mathbf{W}_A$ and $\mathbf{W}_B$ into the optimization and fix others. the optimization problem can be written as:

$$\begin{gathered}\min_{\mathbf{W}_A}\lambda_1\left\|\mathbf{Z}_A-\mathbf{W}_A^{-1}\mathbf{M}_B\right\|_F^2+\lambda_2\left\|\mathbf{W}_A^{-1}\right\|_F^2\\ \min_{\mathbf{W}_B}\lambda_1\left\|\mathbf{Z}_B-\mathbf{W}_B^{-1}\mathbf{M}_A\right\|_F^2+\lambda_2\left\|\mathbf{W}_B^{-1}\right\|_F^2\end{gathered}\tag{8}$$

These are ridge regression problems. According to [20], $\mathbf{W}_A^{-1}$ and $\mathbf{W}_B^{-1}$ are invertible, so the solutions can be obtained as:

$$\begin{aligned}\mathbf{W}_A^{-1}&=\mathbf{Z}_A\mathbf{M}_B^T\left(\mathbf{M}_B\mathbf{M}_B^T+(\lambda_2/\lambda_1)I\right)^{-1}\\ \mathbf{W}_B^{-1}&=\mathbf{Z}_B\mathbf{M}_A^T\left(\mathbf{M}_A\mathbf{M}_A^T+(\lambda_2/\lambda_1)I\right)^{-1}\end{aligned}\tag{9}$$

Updating $\mathbf{P}_A$ and $\mathbf{P}_B$ When others are fixed, the analysis dictionaries $\mathbf{P}_A$ can be updated as follows:

$$\min_{\mathbf{P}_A}\|\mathbf{P}_A\mathbf{X}_A-\mathbf{Z}_A\|_F^2\tag{10}$$

We set the derivative with respect to $\mathbf{P}_A$ to zero, the solution of Eq.(10) can be derived as:

$$\mathbf{P}_A=\mathbf{Z}_A\mathbf{X}_A^T\left(\mathbf{X}_A\mathbf{X}_A^T+\gamma I\right)^{-1}\tag{11}$$

Where γ is a regularization parameter. In our experiments, γ is set to 0.0001. The solution of $\mathbf{P}_B$ is similar to $\mathbf{P}_A$.

Updating $\mathbf{D}_A$ and $\mathbf{D}_B$ With fixed others, we update synthesis dictionary $\mathbf{D}_A$ as below:

$$\min_{\mathbf{D}_A}\|\mathbf{X}_A-\mathbf{D}_A\mathbf{Z}_A\|_F^2\quad s.t.\|d_{A,i}\|_2^2\le 1\tag{12}$$

There are many methods to solve (12). In this paper, we use ADMM algorithm as introduced in [17] to process Eq.(12). The solution of $\mathbf{D}_B$ is similar to $\mathbf{D}_A$.

2.4. Matching

Given the gallery set from camera A and the probe set from camera B, the representation coefficients of the j-th gallery image $p_{A,j}$ and the k-th probe image $p_{B,k}$ are computed with the learned coupled synthesis dictionary as follows:

$$\alpha_{A,j}=\arg\min_{\alpha_{A,j}}\|p_{A,j}-\mathbf{D}_A\alpha_{A,j}\|_F^2+\mu\|\alpha_{A,j}\|_1\tag{13}$$

$$\alpha_{B,k}=\arg\min_{\alpha_{B,k}}\|p_{B,k}-\mathbf{D}_B\alpha_{B,k}\|_F^2+\mu\|\alpha_{B,k}\|_1\tag{14}$$

Where μ is a regularization parameter. Eq.(13) and Eq.(14) can be solved by using the SPAMS toolbox [12]. The modified cosine similarity [16] is employed to compute similarity score between the representation coefficients $\alpha_{A,j}$ and $\alpha_{B,k}$.

3. EXPERIMENTS

3.1. Datasets and Settings

Datasets We evaluated our method on two widely used benchmark datasets: **VIPeR** [21] and **CUHK01** [22]. VIPeR dataset is composed of 632 persons from two non-overlapping camera views, with two images per person. CUHK01 dataset contains 971 persons captured from two non-overlapping camera views. Each person has two images in per camera view. Therefore, there are a total of 3884 images. In our

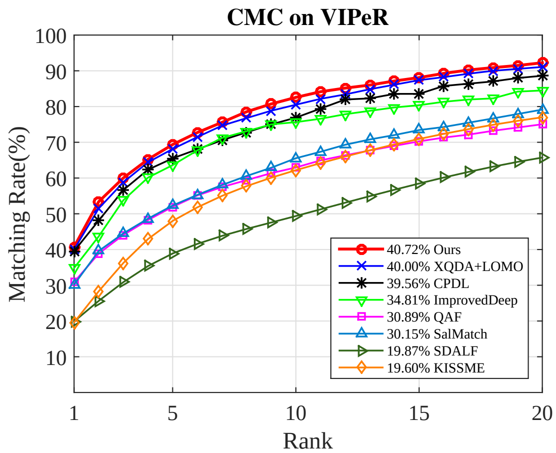

Fig. 2: CMC curve on VIPeR dataset

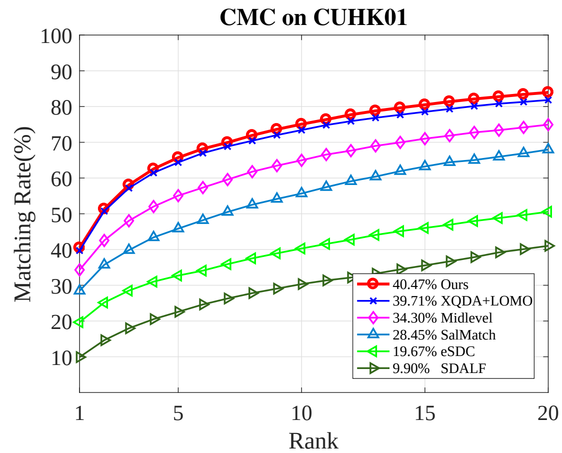

Fig. 3: CMC curve on CUHK01 dataset

experiments, we adopt a single-shot experiment setting. For VIPeR datasets, 316 pairs of images are random selected for training and the rest of image pairs for testing. For CUHK01 datasets, we randomly select one image from per view, so there are 485 for training and 486 for testing. We conduct 10 trials for the averaged results reported. Cumulative Matching Characteristic(CMC) curves are used to evaluate the performance of the proposed method.

Features For feature extraction, we utilize the recently proposed Local Maximal Occurrence (LOMO) features [4].

Parameter Settings To obtain better performance, we set the dimensions of LFDA $m = 800$ for VIPeR and $m = 1000$ for CHUK01, respectively. We set the regularization parameters λ_1=4, λ_2= 0.01 and λ_3= 0.0001. The analysis-synthesis dictionary size d_1 and d_2 are set to 100.

3.2. Results and Comparisons

Fig.2 and Table 1 show the matching results of the compared methods on the VIPeR dataset. From this figure, we can see the proposed method outperforms other approaches [2, 23, 4, 13, 8, 24, 25]. Table 1 summarizes the compared results more clearly at rank-1, 5, 10, 20. It can be seen that our proposed method achieves 40.73% matching rate at rank-1. Although we use LOMO feature as well, our method obtains improvement over the second best result LOMO+XQDA. It is worthy to note that comparing with DPDL which also uses analysis-synthesis dictionary learning method, our method achieves better performance, validating the effectiveness of the proposed method.

Fig.3 and Table 2 show the matching results of the compared methods on the CHUK01 dataset. As compared with other approaches [3, 2, 4, 23, 26], our method achieves higher matching rate at each rank. It is clear that even though the number of test images and the image scenes are different between two datasets (316 on VIPeR and 486 on CUHK01), we all obtain at least 40% rank-1 matching rate, highlighting the robustness and effectiveness of the proposed method in single-shot modality.

Table 1: Top ranked matching rate in (%) on VIPeR

Method	VIPeR(p=316)			
	r=1	r=5	r=10	r=20
KIMSEE [25]	19.60	48.00	62.20	77.00
SDALF [2]	19.87	38.89	49.37	65.73
SalMatch [23]	30.16	52.31	65.54	79.15
QAF [24]	30.89	51.95	62.96	75.05
ImprovedDeep [8]	34.81	63.61	75.63	84.49
CPDL [13]	39.56	65.51	76.90	88.61
XQDA+LOMO [4]	40.00	68.13	80.51	91.08
Ours	**40.73**	**69.37**	**82.56**	**92.25**

Table 2: Top ranked matching rate in (%) on CUHK01

Method	CUHK01(p=486)			
	r=1	r=5	r=10	r=20
SDALF [2]	9.90	22.57	30.33	41.03
eSDC [3]	19.67	32.72	40.29	50.58
SalMatch [23]	28.45	45.85	55.68	67.95
Midlevel [26]	34.30	55.06	64.96	74.94
XQDA+LOMO [4]	39.71	64.36	73.42	81.83
Ours	**40.47**	**65.72**	**75.06**	**83.95**

4. CONCLUSION

In this paper, we propose a coupled analysis-synthesis dictionary learning method, which is capable of enhancing the representation power of the coupled synthesis dictionary. An efficient iterative algorithm is developed for solving the optimization problem in the proposed method. Experimental results on two public person re-identification datasets demonstrate the effectiveness of the proposed method.

5. REFERENCES

[1] S. Gong, M. Cristani, S. Yan, and C. C. Loy, *Person re-identification*, vol. 1, Springer, 2014.

[2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *CVPR*, 2010, pp. 2360–2367.

[3] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *CVPR*, 2013, pp. 3586–3593.

[4] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *CVPR*, 2015, pp. 2197–2206.

[5] J. You, A. Wu, X. Li, and W. S. Zheng, "Top-push video-based person re-identification," in *CVPR*, 2016, pp. 1345–1353.

[6] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *ECCV*, 2014, pp. 1–16.

[7] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, and Y. Y. Tang, "Person re-identification by dual-regularized kiss metric learning," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2726–2738, 2016.

[8] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *CVPR*, 2015, pp. 3908–3916.

[9] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "Joint detection and identification feature learning for person search," *arXiv preprint arXiv:1604.01850*, 2017.

[10] S. Bai, X. Bai, and Q. Tian, "Scalable person re-identification on supervised smoothed manifold," *arXiv preprint arXiv:1703.08359*, 2017.

[11] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian, "Unsupervised cross-dataset transfer learning for person re-identification," in *CVPR*, 2016, pp. 1306–1315.

[12] E. Kodirov, T. Xiang, and S. Gong, "Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification," in *BMVC*, 2015, vol. 3, p. 8.

[13] S. Li, M. Shao, and Y. Fu, "Cross-view projective dictionary learning for person re-identification," in *IJCAI*, 2015, pp. 2155–2161.

[14] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu, "Semi-supervised coupled dictionary learning for person re-identification," in *CVPR*, 2014, pp. 3550–3557.

[15] X. Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, and B. Xu, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," in *CVPR*, 2015, pp. 695–704.

[16] L. An, X. Chen, S. Yang, and B. Bhanu, "Sparse representation matching for person re-identification," *Information Sciences*, vol. 355, pp. 74–89, 2016.

[17] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *NIPS*, 2014, pp. 793–801.

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

[19] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *ICML*, 2006, pp. 905–912.

[20] D. A. Huang and Y. C. Frank Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *ICCV*, 2013, pp. 2496–2503.

[21] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *PETS*, 2007.

[22] W. Li and X. Wang, "Locally aligned feature transforms across views," in *CVPR*, 2013, pp. 3594–3601.

[23] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *ICCV*, 2013, pp. 2528–2535.

[24] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "Query-adaptive late fusion for image search and person re-identification," in *CVPR*, 2015, pp. 1741–1750.

[25] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *CVPR*, 2012, pp. 2288–2295.

[26] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *CVPR*, 2014, pp. 144–151.