Person Re-identification on Heterogeneous Camera Network

Jiaxuan Zhuo[1,3,4], Junyong Zhu[2,3,4], Jianhuang Lai[2,3,4(✉)],
and Xiaohua Xie[2,3,4]

[1] School of Electronics and Information Technology, Sun Yat-sen University,
Guangzhou, China
[2] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
stsljh@mail.sysu.edu.cn
[3] Guangdong Key Laboratory of Information Security Technology,
Guangzhou, China
[4] Key Laboratory of Machine Intelligence and Advanced Computing,
Ministry of Education, Sun Yat-sen University, Guangzhou, China
zhuojx5@mail2.sysu.edu.cn, {zhujuny5,xiexiaoh6}@mail.sysu.edu.cn

Abstract. Person re-identification (re-id) aims at matching person images across multiple surveillance cameras. Currently, most re-id systems highly rely on color cues, which are only effective in good illumination conditions, but fail in low lighting conditions. However, for security issues, it is very important to conduct surveillance in low lighting conditions. To remedy this problem, we propose using depth cameras to perform surveillance in dark places, while using traditional RGB cameras in bright places. Such a heterogeneous camera network brings a challenge to match images across depth and RGB cameras. In this paper, we mine the correlation between two heterogeneous cues (depth and RGB) on both feature-level and transformation-level. As such, depth-based features and RGB-based features are transformed to the same space, which alleviates the problem of cross-modality matching between depth and RGB cameras. Experimental results on two benchmark heterogeneous person re-id datasets show the effectiveness of our method.

Keywords: Person re-identification · Multi-modality matching
Heterogeneous camera network

1 Introduction

Nowadays, the Closed Circuit Television (CCTV) has been widely deployed in many security-sensitive places such as individual house, museum, bank, etc. Because of the economical or privacy issues, there are always non-overlapping regions between different camera views. Therefore, re-identifying pedestrians across different camera views is a critical and fundamental problem for intelligent video surveillance such as cross-camera person searching and tracking. Such a problem is called person re-identification (re-id).

Fig. 1. Illustration of RGB and depth images. Row 1 shows the RGB images in the bright environment. Row 2 shows the RGB images in the dark environment. Row 3 shows the depth images in the dark environment.

Currently, most CCTV systems are based on RGB cameras, and thus the corresponding re-id approaches mainly rely on appearance features. However, in dark environment, appearance features may be unreliable since limited information perceived by RGB cameras. Hence, it is necessary to apply new devices to the dark environment. One alternative solution is to use depth cameras, such as Kinect [1,8,9]. Depth cameras provide depth information and body skeleton joints which are invariant to illumination changes (see Fig. 1). That is, depth cameras remain valid in the dark environment. Depth information is more related to human body shape, and thus beneficial to the re-id in the dark environment [9]. RGB cameras and depth cameras, therefore, form a heterogeneous surveillance network. Previous works in person re-id focus on either RGB camera network [2–7] or depth camera network [8,9], yet none of them addresses the heterogeneous camera network that contains both RGB and depth cameras. In this paper, we focus on how to match pedestrians across depth and RGB cameras in the heterogeneous surveillance network, which has not been studied before.

Along with the line of the traditional person re-id framework [3–6,9], our cross-modality re-id system contains two phases: feature extraction and similarity measurement. Different from [3–6,9], the key idea behind our approach is to mine the correlation between different modalities. In the feature extraction phase, considering that color and texture features [3–7] are not available in depth cameras, we propose to extract body shape information which intrinsically exists in both RGB and depth images. Specifically, for RGB images, we respectively extract two kinds of typical edge gradient features, the classic Histogram of Oriented Gradient (HOG) [7] and the recent proposed Scale Invariant Local Ternary Patterns (SILTP) [5]. Both of them are widely used for shape descriptors. For depth images, we extract the Eigen-depth feature that is recently proposed for Kinect-based person re-id [9]. Note that, both edge gradient feature and Eigen-depth feature describe human body shape and thus can reduce the discrepancy between two kinds of different-modality features. Unfortunately, this correlation is far from a good solution to the cross-modality matching. To address this
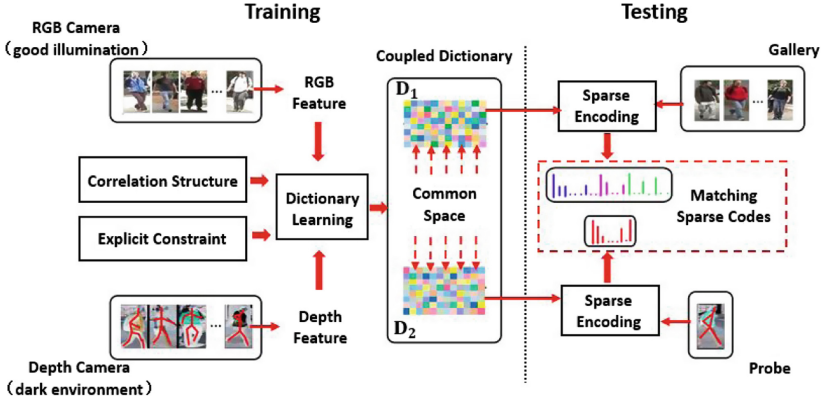
Fig. 2. Overview of our proposed approach. In the training phase, labeled image pairs from RGB and depth cameras are used to jointly learn the discriminative coupled dictionaries optimized by correlation structure and explicit constraint term. In the testing phase, we encode the features of different modalities through the coupled dictionaries as new representations for matching.

problem, we need to mine more correlation between the two modalities. Therefore, we propose a dictionary learning based algorithm that transforms edge gradient feature and Eigen-depth feature into sparse codes that share a common space. In this way, the similarity of edge gradient feature and Eigen-depth feature can be measured with the learned sparse codes. Figure 2 shows the overview of our approach.

In this paper, we identify the dark environment problem in person re-id suffered by unreliable and limited information from RGB cameras, and address this key problem through a novel cross-modality matching approach. To summarize, our contributions include:

- It is a new attempt to the re-id task across depth and RGB modality that we propose a dictionary learning based method to encode different-modality body shape features (edge gradient feature and Eigen-depth feature) into a common space.
- To enforce the discriminability of the learned dictionary pair, we design an explicit constraint term for dictionary learning so that our approach is more discriminative than several contemporary dictionary learning methods.
- Experiments on two heterogeneous person re-id benchmark datasets show the effectiveness of our approach.

2   Proposed Method

2.1   Problem Specification

For the training phase, $F_1 = [f_{11}, f_{21}, \ldots, f_{i1}]$ and $F_2 = [f_{12}, f_{22}, \ldots, f_{i2}]$ denote the gallery and the probe descriptor matrices, respectively, where f_{ij} is the

feature set of the i_{th} training sample. F_1 and F_2 are from two heterogeneous cameras (depth camera and RGB camera) belonging to two different modalities with different dimensions, d_1 and d_2. The goal is to learn the dictionaries $D_1 \in \mathbb{R}^{d_1 \times k}$ and $D_2 \in \mathbb{R}^{d_2 \times k}$ jointly, where k is the size of sparse code. Let $C_1 = [c_{11}, c_{21}, \ldots, c_{i1}]$ and $C_2 = [c_{12}, c_{22}, \ldots, c_{i2}]$ denote the sparse codes of F_1 and F_2, each column of which $c_{ij} \in \mathbb{R}^k$ is the sparse code representing the i_{th} sample.

For the testing phase, feature matrices, $F_G = [f_1^G, f_2^G, \ldots, f_i^G]$ and $F_P = [f_1^P, f_2^P, \ldots, f_i^P]$, extracted from the gallery and the probe, and correspondent sparse codes are $C_G = [c_1^G, c_2^G, \ldots, c_i^G]$ and $C_P = [c_1^P, c_2^P, \ldots, c_i^P]$, respectively.

2.2 Correlative Dictionary Learning

In traditional dictionary learning problem [10], the smaller reconstruction error contributes to a superior dictionary. Hence, we learn a dictionary pair by minimizing two sets of reconstruction errors. Besides, we constrain sparse codes C_1 and C_2 by L_1 regularization similar to sparse representation [10]. To prevent overfitting, we additionally incorporate L_2 regularization for dictionaries and formulate the optimization problem as:

$$\underset{D_1, D_2, C_1, C_2}{\arg\min} \{\|F_1 - D_1 C_1\|_F^2 + \|F_2 - D_2 C_2\|_F^2 + \\ \lambda_C \|C_1\|_1 + \lambda_C \|C_2\|_1 + \lambda_D \|D_1\|_F^2 + \lambda_D \|D_2\|_F^2\} \tag{1}$$

where λ_C and λ_D are regularization parameters to balance the terms.

According to Least Square Semi-Coupled Dictionary Learning (LSSCDL) [11], L_1 regularization on sparse codes is more likely to destroy correlation structure of features, which is suggested to be replaced by L_2 regularization. Many researches [11–14] have proved that L_2 regularization can also play the effect of sparse representation. Therefore, in this paper, we also use L_2 regularization on sparse codes to improve Eq. (1).

Because of the difference between RGB-based and depth-based features, the direct matching results are always unsatisfied. So we capture the correlation between those same persons of cross-modality and penalize those largely different-class scatter. We consider minimizing the Euclidean distance between two sparse codes, namely $\|C_2 - C_1\|_F^2$, to develop the correlation between two dictionaries. Therefore, the objective function is given by:

$$\underset{D_1, D_2, C_1, C_2}{\arg\min} \{\|F_1 - D_1 C_1\|_F^2 + \|F_2 - D_2 C_2\|_F^2 + \lambda \|C_2 - C_1\|_F^2 + \\ \lambda_C \|C_1\|_F^2 + \lambda_C \|C_2\|_F^2 + \lambda_D \|D_1\|_F^2 + \lambda_D \|D_2\|_F^2\} \tag{2}$$

where λ is a positive value which controls the tradeoff between the reconstruction errors and the distance between sparse coding matrices.

In our model, we seek for a discriminative dictionary pair that is able to be discriminative between the same pair and different pairs. We perform this

discriminability by enforcing the constraint on the sparse coefficients corresponding to the learning dictionaries. Let $d_{ii} = c_{i1} - c_{i2}$ denote the Euclidean distance between sparse coefficients corresponding to the gallery and the probe of the same person i, and $d_{ij} = c_{i1} - c_{j2}$ denote the same form of different persons i and j. Specifically, we optimize our model such that the distance between the same person is much smaller than different persons, namely,

$$d_{ii} < d_{ij}, \forall j \neq i, \forall i \tag{3}$$

Thus we optimize the objective function by imposing explicit constraint term:

$$\text{s.t.} \quad \|c_{i1} - c_{i2}\|_2^2 < \|c_{i1} - c_{j2}\|_2^2 \tag{4}$$
$$\forall j \neq i, \forall\, i$$

To simplify optimization, we build the objective function as convex function. Therefore, the constraint term could be modified as:

$$\text{s.t.} \quad \|c_{i1} - c_{i2}\|_2^2 < s_1, \forall\, i \tag{5}$$
$$\|c_{i1} - c_{j2}\|_2^2 < s_2, \forall j \neq i, \forall\, i$$

where s_1 and s_2 are two constants, and $s_1 \ll s_2$. s_1 and s_2 are used to limit the distance between the samples.

In summary, the optimization problem of dictionary learning is described as:

$$\arg\min_{D_1, D_2, C_1, C_2} \sum_{i=1}^{n} \{\|f_{i1} - D_1 c_{i1}\|_2^2 + \|f_{i2} - D_2 c_{i2}\|_2^2 + \lambda\|c_{i2} - c_{i1}\|_2^2$$
$$+ \lambda_C \|c_{i1}\|_2^2 + \lambda_C \|c_{i2}\|_2^2 + \lambda_D \|D_1\|_F^2 + \lambda_D \|D_2\|_F^2\} \tag{6}$$
$$\text{s.t.} \quad \|c_{i1} - c_{i2}\|_2^2 < s_1, \forall\, i$$
$$\|c_{i1} - c_{j2}\|_2^2 < s_2, \forall j \neq i, \forall\, i$$

We employ the alternating optimization algorithm to solve Eq. (6). Specically, we alternatively optimize over D_1, D_2, C_1 and C_2 one at a time, while fixing the other three. Firstly fix D_1, D_2, C_2 and use CVX [20] to optimize a column c_{i1} of C_1. The way to optimize C_2 is similar to C_1. And then we get D_1 and D_2 at gradient algorithm, which are given by

$$D_1 = (F_1 C_1^T)(C_1 C_1^T + \lambda_D I)^{-1} \tag{7}$$

$$D_2 = (F_2 C_2^T)(C_2 C_2^T + \lambda_D I)^{-1} \tag{8}$$

where I is a $k \times k$ identity matrix.

In this way, we alternatively optimize over D_1, D_2, C_1 and C_2 until convergency. The algorithm for training is summarized in Algorithm 1.

Algorithm 1. Correlative dictionary learning

Input: feature matrices F_1 and F_2, size of sparse code k, parameters λ, λ_C, λ_D, $s1$, $s2$
Output: coupled dictionaries D_1 and D_2
Initialize: D_1, D_2, C_1, C_2
while not converge **do**
Step 1: fix D_1, D_2, C_2, update each column of C_1 using CVX by Eq. (6)
Step 2: fix D_1, D_2, C_1, update each column of C_2 using CVX by Eq. (6)
Step 3: fix D_2, C_1, C_1, update D_1 by Eq. (7)
Step 4: fix D_1, C_1, C_2, update D_2 by Eq. (8)
end while

2.3   Person Re-identification by Our Framework

Using the correlative dictionary pair D_1 and D_2, we can obtain the sparse representations of the gallery and the probe. According to Eq. (6), the sparse code coefficients $C_G = [c_1^G, c_2^G, \ldots, c_i^G]$ and $C_P = [c_1^P, c_2^P, \ldots, c_i^P]$ can be respectively obtained by

$$\arg\min_{c_i^G} \left\| f_i^G - D_1 c_i^G \right\|_2^2 + \lambda_G \left\| c_i^G \right\|_2^2, \forall\, i \tag{9}$$

$$\arg\min_{c_i^P} \left\| f_i^P - D_2 c_i^P \right\|_2^2 + \lambda_P \left\| c_i^P \right\|_2^2, \forall\, i \tag{10}$$

where λ_G and λ_P are regularization parameters to balance the terms for the gallery and the probe, respectively.

We use CVX to solve the problems in Eqs. (9) and (10). The algorithm for testing is summarized in Algorithm 2. Finally, the learned sparse codes C_G and C_P are taken as correlative reconstructive features to identity matching by computing the similarity according to the Euclidean distance. In this way, the computational efficiency of identity matching is the same as the standard sparse representation in person re-id [21,22].

Algorithm 2. Sparse representation

Input: feature matrices F_G and F_P, coupled dictionaries D_1 and D_2, parameters λ_G, λ_P
Output: sparse code coefficients C_G and C_P
Initialize: C_G, C_P
Step 1: obtain each column of C_G using CVX by Eq. (9)
Step 2: obtain each column of C_P using CVX by Eq. (10)
Step 3: normalize C_G and C_P

3   Experiment

3.1   Datasets and Features

Datasets. We evaluate our approach on two RGB-D person re-id datasets RGBD-ID [19] and BIWI RGBD-ID [15] collected by Kinect cameras.

BIWI RGBD-ID [15] has three groups, namely "*Training*", "*Still*" and "*Walking*", which respectively contains 50, 28 and 28 humans with different clothing. Each person has 300 frames of RGB images, depth images and skeletons. We use the complete "*Training*" and "*Still*" set and hence there are 78 samples in total. And then we select one frame including RGB and depth for each sample. By convention, we randomly choose about half of the samples, 40 pedestrians for training and the other for testing.

RGBD-ID [19] contains 79 identities with five RGB images, five point clouds and skeletons. We randomly sample approximately half of people (41 identities) in "*Walking1*" for training and the rest for testing because groups "*Walking1*" and "*Walking2*" contain the same person with different frontal views. Only one frame with all information for each person is randomly selected to experiment.

Features. Torso and head are segmented from each image and divided into 6 × 2 rectangular patches by image preprocessing. We obtain integral features of each image by combining local features of each patch. To test the ability of our model adapted to different representations, we consider two kinds of representative edge features, HOG [7] and SILTP [5] as RGB-based features in our experiment. For each depth image, we combine Eigen-depth feature with skeleton information for complete representation as depth-based features [9]. Both HOG and SILTP features capture local region human body shape, as well as depth-based features. Note that RGB-based features and depth-based features belong to different modalities.

3.2   Experiment Settings

Methods for Comparison. To evaluate the effectiveness of our approach, we compare our method with Least Square Semi-Coupled Dictionary Learning (LSSCDL) [11], Canonical Correlation Analysis (CCA) [16]. We also set a baseline which is the multi-modality matching result without any connection between RGB-based and depth-based features for comparison. CCA is a coherent subspace learning algorithm which projects two sets of random variables to the correlated space so as to maximize the correlation between the projected variables in correlated space. LSSCDL is a similar dictionary learning based algorithm, which learns a pair of dictionaries and a mapping function efficiently to investigate the intrinsic relationship between feature patterns. Recently, CCA and LSSCDL have been applied to re-id problem of matching people across disjoint camera views, involving multi-view or multi-modality tasks. CCA and LSSCDL can be used to address the multi-modality matching problem because they can provide connection between uncorrelated variables.

Evaluation Metrics. Recognition rates at selected ranks and the histograms are used to evaluate the performance. The rank-n rate represents the expectation of finding the correct match in the top n matches [17] and rank-1 rate plays an important role to determine the performance of re-id. To ensure fair comparison, the same training and testing samples are used in all methods and the experiments are conducted 10 times to gain the average results.

Parameter Settings. In the following experiments, we set $k = 100$, $\lambda = 0.1$, $\lambda_C = \lambda_D = 0.001$, $\lambda_G = \lambda_P = 0.01$, $s_1 = 0.1$, $s_2 = 100$ for our method. All parameters of other methods are set as suggested in their papers [11,16].

3.3 Experiment Results

Result on BIWI. To prove the universal applicability of our approach, we extract two kinds of typical RGB-based features, HOG and SILTP, to match depth-based features, respectively. Each experiment is carried out in two cases. One is depth-based features for the gallery and RGB-based features for the probe. The other is the reverse. The results are shown in Table 1 and Fig. 3. It can be seen that our method largely outperforms the baseline, which shows the effectiveness of our method to address the multi-modality matching problem. Compared with CCA, our method can establish closer connection than CCA. The main reason is that our method allows screening the vital information and reduces the influence of invalid elements by sparse representation while CCA can not. Our method also outperforms LSSCDL generally, which demonstrates that the explicit constraint term enforces the discriminability of the learning dictionary pair.

Result on RGBD-ID. In RGBD-ID, people's head has be blurred in each RGB images, so the problem becomes more challenging. Following the protocol of experiment on BIWI, we compare the methods in [11,16] using the same feature as that on BIWI in two cases. Table 1 and Fig. 4 show that our method presents the best performance in rank-1 rate. Note that, the margins between the proposed model, CCA, and LSSCDL are small, because the blurred images

Table 1. Recognition rates (%) of cross HOG/SILTP and depth feature on BIWI and RGBD-ID dataset. F_g: Gallery; F_p: Probe; D: Depth; H: HOG; S: SILTP.

	Feature	F_g-D,F_p-H				F_g-H,F_p-D				F_g-D,F_p-S				F_g-S, F_p-D			
	Rank	1	5	10	20	1	5	10	20	1	5	10	20	1	5	10	20
BIWI	RANDOM	2.63	12.11	23.16	50.53	3.15	15.00	26.32	56.58	2.37	12.63	25.53	52.89	2.63	15.53	28.42	55.00
	CCA[13]	6.31	24.21	40.79	64.47	6.31	27.63	40.79	63.68	8.42	26.32	41.58	65.26	6.58	27.37	45.00	72.90
	LSSCDL[11]	8.42	27.11	46.05	70.79	7.11	28.42	41.32	61.32	9.47	24.21	40.26	64.74	7.37	29.47	50.26	74.74
	Ours	11.84	28.42	44.47	70.00	11.32	30.26	48.16	70.26	12.11	26.32	41.58	68.16	9.21	26.32	46.05	71.05
RGBD-ID	RANDOM	2.11	12.63	27.11	55.26	2.63	10.26	23.16	51.05	2.89	13.42	26.32	53.16	1.32	10.26	24.47	51.58
	CCA[13]	7.11	19.74	32.11	56.58	4.74	17.89	30.53	55.53	7.37	17.11	31.05	55.79	6.84	20.79	33.16	58.42
	LSSCDL[11]	7.63	19.47	31.32	57.11	8.16	20.00	34.74	57.37	8.16	18.42	31.32	56.05	7.89	17.11	28.95	53.68
	Ours	11.05	20.00	35.00	57.37	8.94	21.31	36.05	60.79	8.95	19.47	31.32	55.79	8.68	20.79	33.42	60.79

(a) Depth-gallery, HOG-probe

(b) HOG-gallery, Depth-probe

(c) Depth-gallery, SILTP-probe

(d) SILTP-gallery, Depth-probe
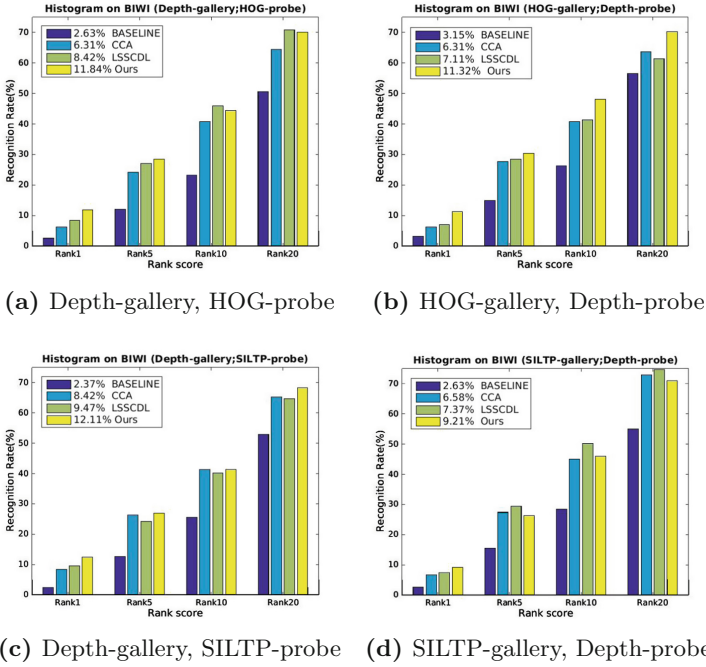
Fig. 3. Histogram and rank-1 rate on BIWI in two cases.

(a) Depth-gallery, HOG-probe

(b) HOG-gallery, Depth-probe

(c) Depth-gallery, SILTP-probe

(d) SILTP-gallery, Depth-probe
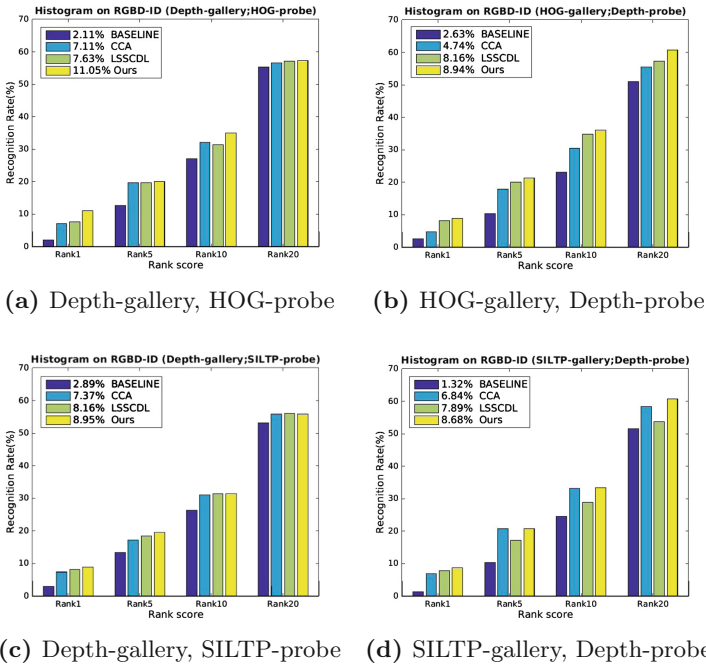
Fig. 4. Histogram and rank-1 rate on RGBD-ID in two cases.

may significantly degrade the discriminability of edge gradient features and thus reduce the correlation between two heterogeneous modalities. With such a weak correlation, the margins between these models cannot be large.

3.4   Effect of Feature Dimensions

We further evaluate the effect of the dimensions of reconstructive features by adjusting the number of the dimensions on BIWI dataset. In particular, we change the dimensions of reconstructive features from 50 to 500 and observe the performance of CCA, LSSCDL and our method. The results in Fig. 5 reflect that (1) reconstructive features with low dimensions outperform those with high dimensions on the whole. The reason may be attributed to the fact that high dimensions are more likely to cause overfitting when the number of training samples is small [18]. (2) The explicit constraint term in Eq. (5) allows to mine more discriminant features, making our method more stable and effective than CCA and LSSCDL in different dimensions.
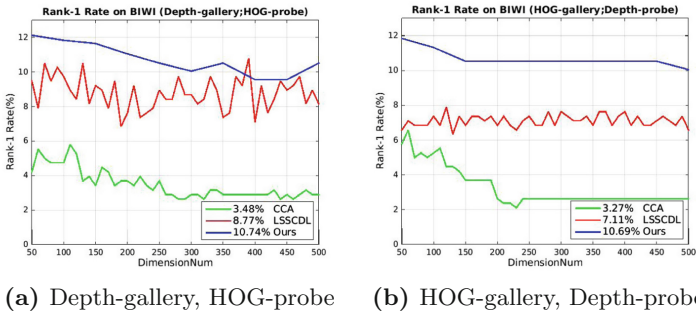

(a) Depth-gallery, HOG-probe   **(b)** HOG-gallery, Depth-probe

Fig. 5. Rank-1 rate on BIWI with variable dimensions of reconstructive feature in two cases and mean average of rank-1 rate.

4   Conclusion

In this paper, we have extended the traditional RGB camera based person re-identification problem to a RGB and depth based cross-modality matching problem. Such a problem is critical when video analysis is needed in heterogeneous camera network. To the best of our knowledge, it is the first attempt for the person re-id to deal with the situation across RGB and depth modality. We have also proposed an effective approach to solve this cross-modality matching problem. It jointly learns coupled dictionaries for RGB and depth camera views. The two views are linked by imposing the two dictionaries to be representative and discriminative. In the testing phase, sparse codes are used for the matching person images across RGB and depth modality. Experimental results on two benchmark

heterogeneous person re-id datasets show the effectiveness and superiority of the proposed approach for multi-modality re-id problem.

In the future, we will carefully integrate correlative dictionary learning into a deep convolutional neural network to jointly learn more robust feature representations and a cross-modality distance metric in an end-to-end way.

References

1. Microsoft Kinect. http://www.xbox.com/en-us/kinect/
2. Chen, S., Guo, C., Lai, J.: Deep ranking for person re-identification via joint representation learning. TIP **25**(5), 2353–2367 (2016)
3. Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person reidentification. PAMI **35**(7), 1622–1634 (2013)
4. Chen, Y., Zheng, W., Lai, J.: Mirror representation for modeling view-specific transform in person re-identification. In: IJCAI (2015)
5. Liao, S., Hu, Y., Zhu, X., Li, S.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR (2015)
6. He, W., Chen, Y., Lai, J.: Cross-view transformation based sparse reconstruction for person re-identification. In: ICPR (2016)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)
8. Albert, H., Alahi, A., Li, F.: Recurrent attention models for depth-based person identification. In: CVPR (2016)
9. Wu, A., Zheng, W., Lai, J.: Robust depth-based person re-identification. TIP **26**(6), 2588–2603 (2017)
10. Lisanti, G., Masi, I., Bagdanov, A.D., et al.: Person re-identification by iterative re-weighted sparse ranking. IEEE Trans. Pattern Anal. Mach. Intell. **37**(8), 1629–1642 (2015)
11. Zhang, Y., Li, B., Lu, H., Irie, A., Ruan, X.: Sample-specific SVM learning for person re-identification. In: CVPR (2016)
12. Shi, Z., Wang, S.: Robust and sparse canonical correlation analysis based L2; p-norm. J. Eng. **1**(1) (2017)
13. Yuan, X., Li, P., Zhang, T.: Gradient hard thresholding pursuit for sparsity-constrained optimization. In: ICML (2014)
14. Shi, X., Guo, Z., Lai, Z., et al.: A framework of joint graph embedding and sparse regression for dimensionality reduction. IEEE Trans. Image Process. **24**(4), 1341–1355 (2015)
15. Munaro, M., Fossati, A., Basso, A., Menegatti, E., Van Gool, L.: One-shot person re-identification with a consumer depth camera. In: Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.) Person Re-Identification. ACVPR, pp. 161–181. Springer, London (2014). https://doi.org/10.1007/978-1-4471-6296-4_8
16. An, L., Kafai, M., Yang, S., Bhanu, B.: Reference-based person reidentification. In: AVSS (2013)
17. Chen, Y., Zheng, W., Lai, J., Yuen, P.: An asymmetric distance model for cross-view feature mapping in person re-identification. IEEE Trans. Circuits Syst. Video Technol. **27**, 1661–1675 (2016)

18. Zhang, L., Xiang, T., Gong, S.: Learning a discriminative null space for person re-identification. In: CVPR (2016)
19. Barbosa, I.B., Cristani, M., Del Bue, A., Bazzani, L., Murino, V.: Re-identification with RGB-D sensors. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012. LNCS, vol. 7583, pp. 433–442. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33863-2_43
20. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V.D., Boyd, S.P., Kimura, H. (eds.) Recent Advances in Learning and Control, pp. 95–110. Springer, London (2008). https://doi.org/10.1007/978-1-84800-155-8_7
21. Jing, X., Zhu, X., Wu, F., et al.: Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In: CVPR (2015)
22. An, L., Chen, X., Yang, S., et al.: Sparse representation matching for person re-identification. Inf. Sci. **355**, 74–89 (2016)