Coupled Feature Spaces Learning with Joint Graph Regularization for Person Re-identification

Peng Bian
College of Mechanical and Material Engineering
North China University of Technology
Beijing, P.R.China
bianpeng@ncut.edu.cn

Yi Jin
School of Computer & Information Technology
Beijing Jiaotong University
Beijing, P.R.China
yjin@bjtu.edu.cn

Luyue Jiang
School of Computer & Information Technology
Beijing Jiaotong University
Beijing, P.R.China
14125177@bjtu.edu.cn

Yidong Li
School of Computer & Information Technology
Beijing Jiaotong University
Beijing, P.R.China
ydli@bjtu.edu.cn

***Abstract*—Re-identification of individuals has already drawn growing attentions due to the increasing intelligent visual surveillance. Human signature is quite different over a network of cameras and most related work devotes to selecting human features without any distinction. To address the problem, we propose a novel coupled feature space learning with joint graph regularization in this paper. The proposed method aims to learn a joint graph regularized common feature space in which two projection matrices can be matched. In the procedure, we use l_{21}-norm to select relevant and discriminative features from coupled space simultaneously. A joint graph regular term enhances the relevance of different photos from the same person. Comparisons results show the superiority and efficiency of our proposed method with performance measured in terms of Cumulative Match Characteristic curves (CMC) on three challenging datasets.**

Keywords—re-identification; coupled space; joint graph

I. Introduction

The task of re-identification can be formalized as the problem of matching person images observed from different cameras. Since some of the cameras are very small and views of the surveillance system might be non-overlapped. Solving the re-identification problem has attracted a rapidly increasing attention from both academic research communities and industrial laboratories in recent years.

There are still large amount of challenges in person re-id such as ⅰ) different camera angles catch the same individual with large variations; ⅱ) the influence of illumination and complicated background; ⅲ) individual-specific long-term structured activities. Moreover, the similarity of the appearance among different individuals becomes higher, leading to more difficulties of re-id.

To address these challenges and explore significant result, some studies [12, 13] have investigated and sought a distance between learning method and reliable low-level feature representation of person views, e.g. Relative Distance Comparison (RDC) model [7], and symmetry-driven accumulation [8].

But it is extremely difficult to compute both distinct and useful local low-level feature from different cameras under complex situations. Existing methods have tried to address the above problem by seeking multiple visual features. Sakrapee. Pk et al. [2] introduced an ensemble of distance function, in which each distance function is learned using a single feature and the final distance is calculated from weighted sum of these distance functions. Nevertheless, the previously mentioned re-id methods typically require enormous labelled target pairs which may not be sufficient in practice.

To bridge the person view variations across cameras, we propose an efficient method called Coupled Feature Spaces Learning with Joint Graph Regularization for person re-identification in this paper. Our method is inspired by the research on cross-model matching that pictures of the same individual from different cameras can be seen as cross-model data. We use two projection matrices to map the photos from the same person into a common feature space. In learning procedure , l21-norm is imposed on the two projection matrices separately, which leads to selected relevant and discriminative features from coupled feature spaces simultaneously, and joint graph regular term is utilized to exploit the structure information. More importantly, pictures of the same individual from different cameras are consistent with the original graph Laplacian by the joint graph regularization.

Main contributions of our paper can be summarized as follows:

1) We propose a joint graph regularized coupled feature spaces learning method to improve re-id accuracy. The pictures

This work was supported by the National Natural Science Foundation of China (No. 51505004, 61403024), the Beijing Natural Science Foundation (No.4163075, 4162048) and the Fundamental Research Funds for the Central Universities (No. 2015RC086).

from different cameras are complementary to each other and have the local coordinate by the joint graph regularization item, mapping the data with locality-constrained into the coupled feature spaces in which cross cameras data can be performed.

2) An iterative algorithm is presented to efficiently generate the two projection matrices. The problem of training can be simplified to two linear system problems in iteration. In addition, we propose the optimization method.

The rest of the paper is organized as follows: a brief view of related works is presented in Section 2. Section 3 introduces the proposed algorithm and its kernel extension. Experimental results on iLIDS, VIPeR and PRID datasets are presented in Section 4. Finally, the concluding remarks in Section 5.

II. RELATED WORK

Since the person re-identification is considered as an important problem in some real applications, various approaches have been proposed to deal with this problem. The main difficulty of person re-id focus on the variations in human appearances from different camera views, and distance learning-based person re-identification [13] shows significant improvement in performance. Liu et al. [14] introduced a semi-supervised coupled LCC dictionary learning method to bridge the human appearance variations across cameras by using local coordinate coding sparse representation and semi-supervised coupled dictionary learning. This approach avoids treating all features indiscriminately and jointly learning the coupled dictionaries from both labeled and unlabeled data. Wang et al. [1] using a coupled feature space for distance learning, calculated the distance across different modalities in the two related spaces.

As Yu et al. suggested in [15] that under certain assumptions locality is more essential than sparsity. Local feature matching-based person re-identification matches the carefully designed local feature. Zhai et al. [3] proposed the heterogeneous metric learning method of using joint graph regularization item to bridge the different modal data , and the item can implement the locality-constraint for cross-model matching.

Feature representation is one of the most important parts of person re-id systems. Robust and discriminative features can receive a better result, and [16] introduced some more about feature representations. The visual features which applied in our re-id approach are SIFT and HOG, and Scale-invariant feature transform (SIFT) has attracted considerable research attention due to its invariance to scaling, orientation and illumination changes. Histograms of Oriented Gradients (HOG) captures edge or gradient structure that is the main characteristic of local shape, and it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations.

III. OUR APPROACH

In this section, we firstly propose a regularization framework for the person re-id, which can be formulated as a minimization problem. Secondly an iterative algorithm is proposed to optimize the minimization problem.

A. Problem Definition

Notations. Let $N \in R^{m\times n}$. For matrix N, $N^{(i)}$ and N_j are defined as its i-th row and the j-th column respectively. The Frobenius norm of the matrix is:

$$\|N\|_F = \sqrt{\sum_{i=1}^{m} \|N^{(i)}\|_2^2} \tag{1}$$

$\|N\|_{21}$ is the sum of the l2-norm of the rows of :

$$\|N\|_{21} = \sum_{i=1}^{m} \|N^{(i)}\|_2 \tag{2}$$

We first overview the framework of traditional distance metric learning, given two objects X_i and X_j, the purpose of the method is to learn an optimal metric:

$$d(x_i, x_j) = \sqrt{(U^T x_i - U^T x_j)^T (U^T x_i - U^T x_j)} \tag{3}$$

We suggest that the pictures of the same individual from different cameras share the same label Y which can be the corresponding contact between the two category objects. The distance metric can be converted into:

$$d(x_i, x_j) = \sqrt{(U_a^T x_i - Y)^T (Y - U_a^T x_i)} + \sqrt{(U_b^T x_j - Y)^T (Y - U_b^T x_j)} \tag{4}$$

We aim to learn the two parameter matrices Ua and Ub from training the labelled dataset.

Formulating the regularization framework for distance metric learning as follows:

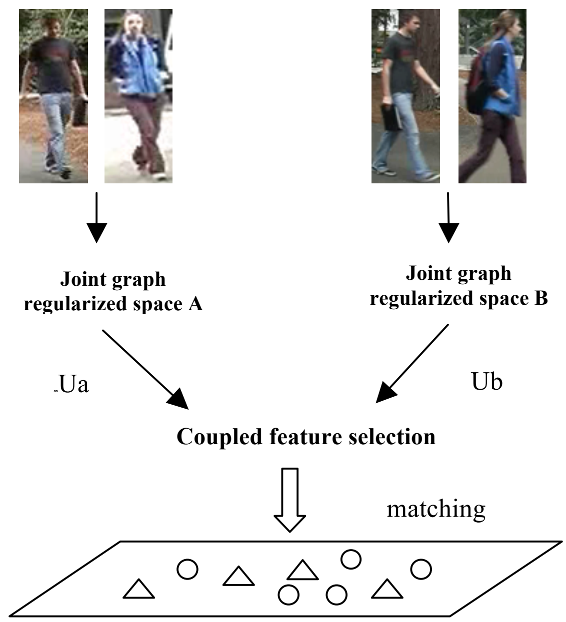


Fig.1. The overview of the proposed method, the joint graph regularized space A and space B are projected to the matching space by the projection matrices Ua and Ub, l_{21}-norm is used for coupled feature selection.

$$\min_{U_a,U_b} f(U_a, U_b) + \lambda_1 g(U_a, U_b) + \lambda_2 w(U_a, U_b) + \lambda_3 r(U_a, U_b) \quad (5)$$

B. Coupled Feature Spaces Learning

Suppose that we have two sets of pictures from two different cameras: $X_a = \{ x_1^a, x_2^a, x_3^a, \dots, x_n^a \} \in R^{d\times n}$, $X_b = \{ x_1^b, x_2^b, x_3^b, \dots, x_n^b \} \in R^{d\times n}$, X_a and X_b represent the features extracted from pictures captured from camera A and camera B respectively, n represents the number of category person, d represents the dimension. Each pair $\{ x_i^a, x_i^b \}$ represents the multi-camera photos of the same pedestrian. Let $Y = [y_1, y_2, y_3, \dots, y_n]^T \in R^{n\times c}$ be the class label matrix, c is the number of the classes. Our model aims to learn two projection matrices to map the data of the joint graph regularized coupled spaces into a common space defined by class labels. In order to solve the problem of measuring the relevance of multi-camera data from same person, two Frobenius norms are used to be the loss function in formula (5). We perform l_{21}-norm on the projection matrices for coupled feature selection and use the trace norm for scale regularization and a regularization norm of joint graph which we briefly discuss below.

The formula can be written as:

$$\min_{U_a,U_b} \frac{1}{2}(\|X_a^T U_a - Y\|_F^2 + \|X_b^T U_b - Y\|_F^2) + \lambda_1 g(U_a, U_b) + \lambda_2(\|U_a\|_{21} + \|U_b\|_{21}) + \lambda_3 \|X_a^T U_a X_b^T U_b\|_* \quad (6)$$

Where U_a and U_b are the projection matrices for the coupled space respectively, $\bar{L}$ is the normalized graph Laplacian.

C. The Joint Graph Regularization

The joint graph regularization item is used to learn the transformation consistent with the locality constraints in both sides of multi-camera data. For the different cameras capture different views, we define a joint undirected graph, G = (V, W) on the datasets. is the element of locality matrix W and W = means the locality between the i-th object and the j-th object. We adopted the label information to construct the symmetric locality matrix and used Locality Preserving Indexing (LPI) [6] method to obtain W.

$$w_{ij} = \begin{cases} \frac{1}{n_l}, \text{if } x_i \text{ and } x_j \text{ both belong to the same label;} \\ 0, \text{otherwise.} \end{cases} \quad (7)$$

The normalized graph Laplacian L is defined as:

$$\bar{L} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad (8)$$

where I is an (2n) × (2n) identity matrix and D is an (2n) × (2n) diagonal matrix with $d_{ii} = \sum_j w_{ij}$. $\bar{L}$ is symmetric and positive semidefinite. We define:

$$O = \left(U_a^T X_a \ U_b^T X_b\right), \quad \bar{L} = \begin{bmatrix} \bar{L}_a & \bar{L}_{ab} \\ \bar{L}_{ba} & \bar{L}_b \end{bmatrix} \quad (9)$$

O represents for the learned metric space. Since:

$$tr\begin{bmatrix} A & B \\ C & D \end{bmatrix} = tr(A) + tr(D) \quad (10)$$

Based on this we formulate the regularization as follows:

$$\begin{aligned} g(U_a, U_b) &= \frac{1}{2}\mathrm{tr}(O\bar{L}O^T) \\ &= \frac{1}{2}\mathrm{tr}(O\bar{L}_a O^T) + \frac{1}{2}\mathrm{tr}(O\bar{L}_b O^T) \\ &= \frac{1}{2}\mathrm{tr}(U_a^T X_a \bar{L}_a U_b^T X_b) + \frac{1}{2} tr(U_b^T X_b \bar{L}_b U_a^T X_a)^T \end{aligned} \quad (11)$$

where tr(X) is the trace of a matrix X and the regularization $g(U_a, U_b)$ penalizes large change of the spaces.

D. Iterative Solution

In order to solve the minimization problem in the objective formula which involves l_{21}-norm and two trace norms, it is difficult to solve directly so we adopt the solution method in [1]. The l_{21}-norm can be converted into:

$$w(U_a, U_b) = \mathrm{tr}(U_a^T P U_a) + tr(U_b^T Q U_b) \quad (12)$$

where P=Diag(p) and Q=Diag(q), p and q are auxiliary vectors of two l_{21}-norms. ε is the smoothing term which is usually set to be a small constant value.

$$\begin{cases} p_i = 1 \Big/ 2\sqrt{\|u_a^i\|_2^2 + \varepsilon} \\ q_i = 1 \Big/ 2\sqrt{\|u_b^i\|_2^2 + \varepsilon} \end{cases} \quad (13)$$

The trace norm of the formula is equal to:

$$\|X_a^T U_a X_b^T U_b\|_* = \frac{1}{2} tr([X_a^T U_a X_b^T U_b]^T S^{-1} [X_a^T U_a X_b^T U_b]) + \frac{1}{2} tr(S) \quad (14)$$

Using method [] to calculate the infimum over S:

$$S = (X_a^T U_a U_a^T X_a + X_b^T U_b U_b^T X_b + \mu I)^{\frac{1}{2}} \quad (15)$$

Given S and calculating the graph norms Lx and Ly, optimizing the objective function (6) over Ua and Ub respectively is equal to optimizing the following two linear problems:

$$\begin{cases} \min_{U_a,U_b} \frac{1}{2}(\|X_a^T U_a - Y\|_F^2) + \frac{\lambda_1}{2}\mathrm{tr}(U_a^T X_a \bar{L}_x X_a^T U_a) + \\ \lambda_2 tr(U_a^T P U_a) + \frac{\lambda_3}{2} tr(U_a^T X_a S^{-1} X_a^T U_a) \\ \min_{U_b} \frac{1}{2}(\|X_b^T U_b - Y\|_F^2) + \frac{\lambda_1}{2}\mathrm{tr}(U_b^T X_b \bar{L}_x X_b^T U_b) + \\ \lambda_2 tr(U_b^T P U_b) + \frac{\lambda_3}{2} tr(U_b^T X_b S^{-1} X_b^T U_b) \end{cases} \tag{16}$$

IV. EXPERIMENTS

In this section, we evaluate our Joint Graph Regularized Coupled Feature Spaces approach with state-of-the-art person re-identification method on three publicly available VIPeR[9], PRID[11] and iLIDS[10] datasets, all of the sets are used for single-shot evaluation. The following describes the details of the experimental setups and the results.

Datasets Several challenging benchmark data sets can be used for person re-identification. In this paper, we use three commonly used data sets (VIPeR, PRID 2011, iLIDS) for training and testing. The VIPeR set [9] is composed of 1264 images of 632 individuals, from two cameras with different viewpoints. The PRID set has 770 images of 386 pedestrians, with 2 images of 128×48 pixels per individual. The iLIDS set contains 476 images of 119 individuals, and the images often have severe occlusions caused by people and luggage, since this dataset was collected at an airport.

Experimental setting In our experiments, we adopted a single-shot experiment setting. All the datasets were separated into two subsets based on the camera numbers. The images from camera A are used as the probe images, while the images

---

Algorithm 1: Iterative Algorithm for Learning Joint Graph Regularized Coupled feature spaces

---

Require: Training data sets X_a and X_b, label information sets Y.

Result: Projection matrices U_a and U_b.

Set t = 0. Initialize U_a and U_b as zero matrix.

repeat

1. Compute W_x, W_y and V Diag(s_k)V^T as the eigenvalue.

2. Compute $\bar{L}_x$ and $\bar{L}_y$ according to (8).

3. Set $S^{-1} = V\ Diag(\frac{1}{s_k + \nu})\, V^T$

4. Compute U_a^t and U_b^t by solving the two linear system problems.

5. t = t +1

until converges

Algorithm 1 summarizes the minimization problem, step one constructs the graph and corresponds to the trace norm, step two computes the graph norm, step three and four are used to calculate the projection matrices.

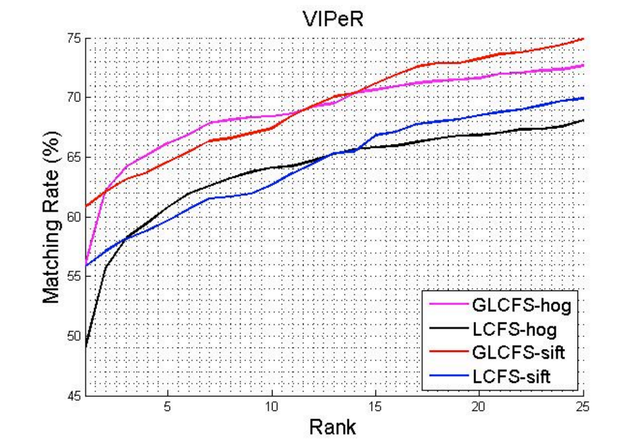
In camera B are used as the gallery images. The mapping evaluation on these datasets is repeated 10 times.

For each comparison between our method and the baseline method LCFS, we extracted two different features from all the datasets, while making the two subsets A and B randomly divided into training part and testing part. This partition was repeated 10 times.

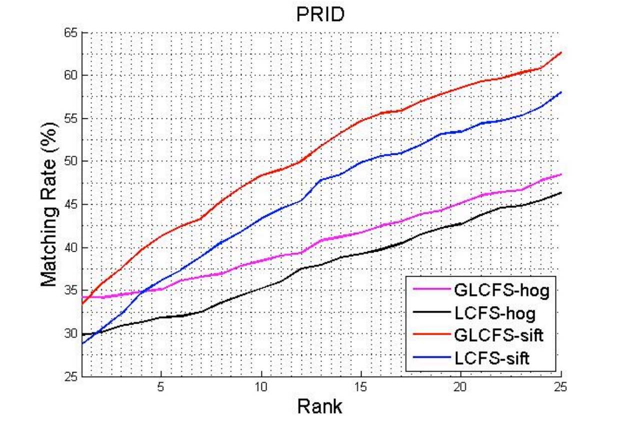
Comparison with state-of-the-art result Our approach is compared with LCFS algorithm in three data sets with two kinds of features (HOG and SIFT). We report the widely used accumulated Cumulative Match Characteristic (CMC) performance curves for easy comparison. In addition, we also report the average rank-1 result of each comparison.

V. CONCLUSIONS

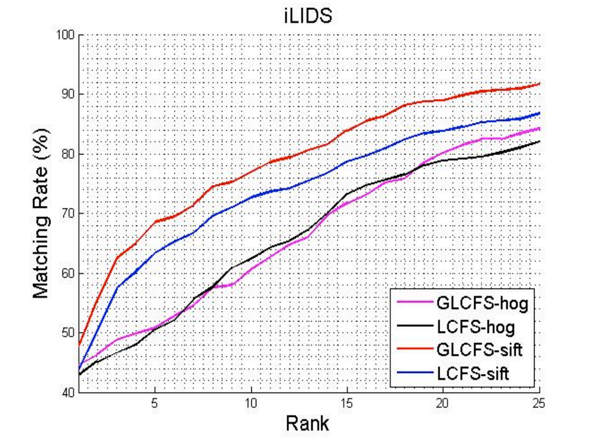
In this paper, we have proposed a Joint Graph Regularized coupled feature spaces (GCFS) method for person re-identification. In the proposed method, we exploit two coupled subspaces which have been graph regularized for multi-camera images identification, and the joint graph regular norm makes the subspaces complementary to each other. Experiments on three different person re-identification datasets demonstrate the effectiveness and generalization of the proposed new method.



(a)



(b)



(c)

Fig. 2. CMC performance for VIPeR, PRID and iLIDS data sets, the higher the recognition rate, the better the performance.

TABLE I
RANK-1 RECOGNITION RATE OF EXISTING BEST REPORTED RESULTS AND OUR RESULTS.

DATA SET	INDIVIDUALS		PREV.BEST	OURS
	train	Test		
VIPeR	316	316	43.4%[17]	46.7%
PRID2011	100	100	17.9%[2]	23.8%
iLIDS	59	60	40.3%[18]	45%

REFERENCES

[1] Kaiye Wang and Ran He. Learning Coupled Feature Spaces for Cross-modal Matching. In ICCV, pages 2088-2095, 2013.

[2] Sakrapee Paisitkriangkrai and Chunhua Shen. Learning to rank in person re-identification with metric ensembles. In Computer Vision and Pattern Recognition (CVPR), 2015.

[3] Xiaohua Zhai, Yuxin Peng, Jianguo Xiao. Heterogeneous Metric Learning with Joint Graph Regularization for Cross-Media Retrieval. In Proceedings of the Twenty-Seventh AAAI Conference on Artificial.

[4] Z. Harchaoui, M. Douze, M. Paulin, M. Dudik, and J. Malick. Large-Scale Image Classification with Trace-norm Regularization. In CVPR, pages 3386–3393, 2012.

[5] Q. Gu, Z. Li, and J. Han. Joint Feature Selection and Subspace Learning. In IJCAI, pages 1294–1299, 2011.

[6] Deng Cai, Xiaofei He, Jiawei Han. Document Clustering Using Locality Preserving Indexing. In IEEE Transactions on, 17(12):1031–1045, 2005.

[7] W. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. IEEE Trans. Pattern Anal. Mach. Intell., 35(3):653-668, 2013.

[8] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. Proc. CVPR, 2010.

[9] Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Computer Vision–ECCV 2008, pp. 262–275. Springer (2008)

[10] Zheng, W.S., Gong, S., Xiang, T.: Associating groups of people. In: BMVC (2009)

[11] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In Proc. Scandinavian Conf. on Image Anal., 2011.

[12] Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 3318–3325. IEEE (2013)

[13] M. Dikmen, E. Akbas, T. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. Proc. ACCV, 2010.

[14] Xiao Liu, Mingli Song, Dacheng Tao, Xingchen Zhou, Chun Chen and Jiajun Bu. Semi-Supervised Coupled Dictionary Learning for Person Re-identification. In Computer Vision and Pattern Recognition (CVPR), 2014.

[15] Kai Yu, Tong Zhang and Yihong Gong. Nonlinear Learning using Local Coordinate Coding. In Neural Information Processing Systems 22 (NIPS 2009).

[16] S. Gong, M. Cristani, S. Yan, and C. C. Loy. Person Re-Identification. Springer, 2014.

[17] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In Proc. IEEE Conf. Comp.Vis. Patt. Recogn., 2014.

[18] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person reidentification using kernel-based metric learning methods. In Proc. Eur. Conf. Comp. Vis., 2014.