SELF-PACED LEAST SQUARE SEMI-COUPLED DICTIONARY LEARNING FOR PERSON RE-IDENTIFICATION

Wei Xu, Haoyuan Chi, Lei Zhou, Xiaolin Huang, Jie Yang

Institution of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China

ABSTRACT

Person re-identification aims to match people across disjoint camera views. It has been reported that Least Square Semi-Coupled Dictionary Learning (LSSCDL) based sample-specific SVM learning framework has obtained the state of the art performance. However, the objective function of the LSSCDL, the algorithm of learning the pairs (feature, weight) dictionaries and the mapping function between feature space and weight space, is non-convex, which usually result in suboptimal solutions with the bad local minima of the objective function. To tackle with this constraint, we present Self-Paced Least Square Semi-Coupled Dictionary Learning (SLSSCDL) algorithm, which is inspired by previous works on self-paced learning, a framework able to improve the accuracy of conventional learning models by presenting the training data in a meaningful order to get a better local minima, i.e. easy samples are provided first. In addition, a graph based regularization term is also introduced to preserve the local similarities in each space. Experimental results show that SLSSCDL gains competitive performance on two challenging datasets.

Index Terms— Self-Paced Learning, samplespecific SVM, person re-identification

1. INTRODUCTION

Person re-identification, a fundamental task in multicamera surveillance system, is to recognize a person of interest over different camera views at different locations based on image appearance. Although person re-identification has received increasing attention in recent years, significant variations on the poses, illumination, viewpoints and appearance for the observed person make this problem pretty challenging.

In general, previous person re-identification solutions [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] generally belong to one of two groups: the invariant features schemes and the metric learning methods.

The first group aims to develop robust feature representations which are discriminative for identity, such as Symmetry-Driven Accumulation of Local Features (SDALF) [1], Covariance descriptor based on Bio-inspired Features (gBiCov) [2], and Local Maximal Occurrence (LOMO) [3]. Then common distance metrics or several discriminative classifiers are directly adopted to these features for matching.

The second group often seeks for a metric in which the examples of the same people are close and those of the different people are far. Among them Probabilistic Relative Distance Comparison (PRDC) [8], Multi-view implicit transfer (MICT) [10], Information Theoretic Metric Learning (ITML) [4], Keep It Simple and Straightforward Metric Learning (KISSME) [5] and Cross-view Quadratic Discriminant Analysis (XQDA) [3], have achieved impressive results.

Schemes for improving matching models are the main focus of this paper. Recently Zhang [7] proposes a LSSCDL based sample-specific SVM learning framework for person re-identification, where a sample-specific SVM is learned for each pedestrian to seek the optimal match. Then it adapts LSSCDL to learn a dictionary pair of feature space and weight space and a mapping function simultaneously, through which the weight parameters of a new sample can be easily inferred by its feature patterns.

However, a common issue with dictionary learning is that the learning algorithms usually find sub-optimal solutions corresponding to bad local minima of the objective function. To address such problem, we propose a self-paced joint learning framework, SelfPaced Least Square Semi-Coupled Dictionary Learning (SLSSCDL), by sorting the training data in a meaningful order to obtain a better local minima. In addition, a graph based regularization term is introduced to preserve the relationship of the local similarities.

Figure 1 illustrates the idea of our proposed method. SLSSCDL first learns with those samples that have a good trade-off between reconstruction error and code mapping matching. Intuitively, our method first selects the sample associated with the red circle, then with the yellow triangle and finally with the blue square. To verify our proposition, we develop a novel coupled dictionary learning framework to learn the dictionaries, codes and mappings and simultaneously infer the optimal sample order. In the end, we evaluate the effectiveness of our framework by conducting experimental

This work is partly supported by National Science Foundation, China (No: 615723156151101179), and 973 PlanChina (No. 2015CB856004). . Wei Xu(xuronaweizida@gmail.com) and Jie Yang(jieyang@sjtu.edu.cn) are the corresponding authors.

evaluations on two challenging datasets.

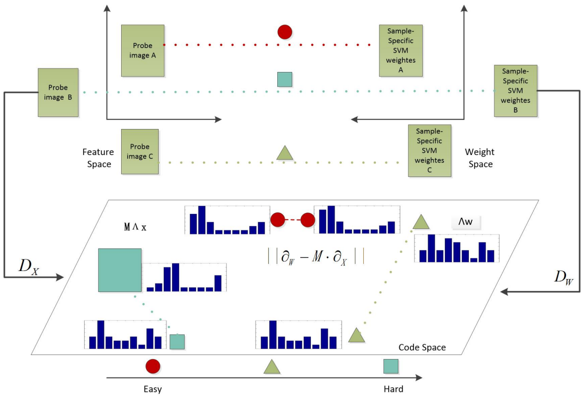
Probe image A
Sample-Specific SVM weightes A
Probe image B
Sample-Specific SVM weightes B
Feature Space
Probe image C
Sample-Specific SVM weightes C
Weight Space
D_X
D_W
M Λ x
Λw
$||\partial_W - M \cdot \partial_X||$
Code Space
Easy
Hard

Fig. 1. Application of the SLSSCDL method to person re-identification. The right side shape size represents sparse codes of dictionaries D_W for Sample-Specific SVM weights and the left side represents the sparse codes of dictionaries D_X for the probe images multiple with the mapping function M between the sparse code space of feature and weight. Intuitively, our method first selects the sample associated with the red circle, then with the yellow triangle and finally with the blue square. Details are given in main text.

2. THE PROPSED ALGORITHM

In this section we present our SLSSCDL approach. We first introduce the sample-specific SVM learning for person re-identification [7]; then Self-paced learning is revisited [11]; finally we describe the proposed SLSSCDL method along with the algorithm we devised to solve the associated optimization problem.

2.1. Sample-Specific SVM Learning for Person Reidentification

Motivated by the insight that different matching functions should be designed for different individuals, [7] formulates the person re-identification problem into a binary classification problem and learn a classifier specifically for each pedestrian. It consists of 3 steps: Sample-Specific SVM Learning, LSSCDL [7] and Pedestrian Matching. A short description for each step is given as follows:

Step 1: Sample-Specific SVM Learning Given a probe image x_i^p, it learn a sample-specific classifier on the probe-gallery set $\{(x_i^p, x_1^g), \ldots, (x_i^p, x_N^g)\}$such that

$$F_i(x_i^p, x_j^g) = \begin{cases} \geq 0, & y_j^p = 1 \\ < 0, & y_j^p = -1. \end{cases} \tag{1}$$

with the form

$$F_i(x_i^p, x_j^g) = w_i^p * \phi(x_i^p, x_j^g) + b_i \tag{2}$$

where w_i^p denotes the weight vector of x_i^p and b_i is the bias, $\phi(x_i^p, x_j^g)$ is defined as a feature map of the image pair.

Step 2: Least Square Semi-Coupled Dictionary Learning LSSCDL [7] learns a pair of dictionaries and a mapping function, where the two dictionaries respectively depict the intrinsic structures of the feature space and weight space, and the mapping function characterizes the relationship between the two spaces. More details will be illustrated in section 2.3.1.

Step 3: Pedestrian Matching Given a test probe image x_t^p, the corresponding weight vectorw_t^p can be derived with the learned dictionary pair D_X, D_W and mapping matrix M. We compute the matching score of the test prob-gallery image pair with (2).

2.2. Self-Paced Learning Revisit

Given a training datasets $D = \{(x_i, y_i)\}_{i=1}^n$, in which (x_i, y_i) denotes the observed sample and its label, then let $L(y_i, g(x_i, w))$ be the loss function, w is the model parameter of the decision function g. Generally, the objective function of self-paced learning [7,9, 10] is expressed as:

$$\begin{aligned} \min_{w,v} \quad & E(w, v; \lambda) = \sum_{i=1}^{n} v_i L(y_i, g(x_i, w)) + f(v, \lambda), \\ \text{s.t.} \quad & v \in [0, 1]^n, \end{aligned} \tag{3}$$

where λ is the age parameter for controlling the learning rate, and $f(v, \lambda)$ is the self-paced regularizer. This strategy has proved helpful in alleviating the local optimal problem in non-convex optimization [12, 13].

2.3. Model Formulation

2.3.1. Least Square Semi-Coupled Dictionary Learning

Given the training probe set $X^P = (X_1^P, X_2^P, \ldots, X_N^P) \in R^{d \times N}$ with each column representing a probe image, and the corresponding learned weight set W^p, we denote $D_X D_W M$ to be the feature dictionary, the weight dictionary, and the mapping matrix, respectively. And then LSSCDL [8] optimizes the dictionaries and mapping function jointly with the following formulation:

$$\begin{aligned} \min_{D_X, D_W, M} & E(D_X, D_W, M) = \left\|X^P - D_X \Lambda_X\right\|_F^2 \\ & + \left\|W^P - D_W \Lambda_W\right\|_F^2 + \lambda \|\Lambda_W - M\Lambda_X\|_F^2 + \lambda_\Lambda \|\Lambda_X\|_F^2 \\ & + \lambda_\Lambda \|\Lambda_w\|_F^2 + \lambda_M \|M\|_F^2 + \lambda_D \|D_X\|_F^2 + \lambda_D \|D_W\|_F^2 \end{aligned} \tag{4}$$

where λ,λ_Λ,λ_M,λ_D are regularization parameters to balance the terms in the objective function and Λ_X, Λ_W denote the coding coefficients

2.3.2. *Locality preserving*

Relationship among the samples is very useful when learning new representations [14, 15]. To embed this prior knowledge into dictionary learning, we propose to use the original set of features to create an undirected proximity graph. The weights of this graph can be computed with the Gaussian kernel. The weights are then used to create the Laplacian matrix L. Therefore two regularization terms $Tr(\Lambda_X^T L_X \Lambda_X)$,$Tr(\Lambda_W^T L_W \Lambda_W)$ are then added to the dictionary problem (3).

2.3.3. *Self-Paced Least Square Semi-Coupled Dictionary Learning*

Associating local similarity preservation with self-paced learning together, we obtain the overall objective function for SLSSCDL:

$$\begin{aligned} &\min_{D_X, D_W, M, \Lambda_X, \Lambda_W, V} E(D_X, D_W, M, \Lambda_X, \Lambda_W, V) \\ &= \left\|(X^P - D_X \Lambda_X)V\right\|_F^2 + \left\|(W^P - D_W \Lambda_W)V\right\|_F^2 \\ &+ \lambda\|(\Lambda_W - M\Lambda_X)V\|_F^2 + \lambda_\Lambda\|\Lambda_X\|_F^2 + \lambda_\Lambda\|\Lambda_w\|_F^2 \\ &+ \lambda_M\|M\|_F^2 + \lambda_D\|D_X\|_F^2 + \lambda_D\|D_W\|_F^2 \\ &+ \lambda_S(Tr(\Lambda_X^T L_X \Lambda_X) + Tr(\Lambda_W^T L_W \Lambda_W)) + f(V, k) \end{aligned} \tag{5}$$

where,λ_S is the regularization parameters of local smoothness, V is the diagonal matrix with the sample weights $V_i \in [0, 1]$ on the diagonal of V, respectively. Therefore SLSSCDL can assess the learning easiness of each sample, not only from the reconstruction error, but also from the correspondence mapping between feature space and weight space.

2.3.4. *Optimization Algorithm*

As the optimization problem for SLSSCDL is not jointly convex, similar as [7, 11, 16], we propose an alternating optimization approach to solve (4) in four steps: codes, dictionaries, mappings and pacing variables.

Solving for codes Λ_X, Λ_W:
Fix $D_X, D_W, M, \Lambda_W, V$,let $\frac{\partial E}{\partial \Lambda_X} = 0$, we have

$$\begin{aligned} \Lambda_X = (&D_X^T D_X V V^T + \lambda M^T M V V^T + \lambda_\Lambda I) \\ &+ 0.5\lambda_S(L_X + L_X^T))^{-1}(D_X^T X^P V V^T + \lambda M^T \Lambda_W V V^T) \end{aligned} \tag{6}$$

Fix $D_X, D_W, M, \Lambda_X, V$,let $\frac{\partial E}{\partial \Lambda_W} = 0$, we have

$$\begin{aligned} \Lambda_W = (&D_W^T D_W V V^T + (\lambda V V^T + \lambda_\Lambda)I) \\ &+ 0.5\lambda_S(L_W + L_W^T))^{-1}(D_W^T W^P V V^T + \lambda M \Lambda_X V V^T) \end{aligned} \tag{7}$$

Solving for dictionaries D_X, D_W:
Fix $\Lambda_X, \Lambda_W, M, V$,let $\frac{\partial E}{\partial D_X} = 0$ and $\frac{\partial E}{\partial D_W} = 0$ we get

$$D_X = X^P V V^T \Lambda_X^T (\Lambda_X V V^T \Lambda_X^T + \lambda_D I)^{-1} \tag{8}$$

```
Algorithm 1 The Optimization of SLSSCDL
input: probe image matrix $X^P$, weight matrix $W^P$, parameters $\lambda, \lambda_\Lambda, \lambda_M, \lambda_D, \lambda_S$
Output: feature dictionary $D_X$ , weight dictionary $D_W$, mapping matrix $M$

    Initialize: $D_X, D_W, \Lambda_X, \Lambda_W$ and $M$
    Repeat:
    1: Fix $D_X, D_W, \Lambda_W, M$ and $V$, update $\Lambda_X$ by (5)
    2: Fix $D_X, D_W, \Lambda_X, M$ and $V$, update $\Lambda_W$ by (6)
    3: Fix $\Lambda_W, \Lambda_X, M$ and $V$, update $D_X, D_W$ by (7) and (8)
    4: Fix $D_X, D_W, \Lambda_W, \Lambda_X$ and $V$, update $M$ by (9)
    4: Fix $D_X, D_W, \Lambda_W, \Lambda_X$ and $M$, update $V$ by (10)
    5: Increase $k$ by $\mu$ to enlarge the training set;
    Until: convergence
```

$$D_W = W^P V V^T \Lambda_W^T (\Lambda_W V V^T \Lambda_W^T + \lambda_D I)^{-1} \tag{9}$$

Solving for mapping M:
Fix $D_X, D_W, \Lambda_X, \Lambda_W, M, V$,let $\frac{\partial E}{\partial M} = 0$, we have

$$M = \Lambda_W V V^T \Lambda_X^T (\Lambda_X V V^T \Lambda_X^T + \frac{\Lambda_M}{\lambda} I)^{-1} \tag{10}$$

Solving for pacing variables V:
Fix $D_X, D_W, \Lambda_X, \Lambda_W, M$, we obtain V

$$V_i = \begin{cases} 1, & if e_n + \lambda_s f_n < k \\ 0, & otherwise. \end{cases} \tag{11}$$

where $e_n = \left\|(X_n^P - D_X \alpha_n^X)\right\|_F^2 + \left\|(W_n - D_X \alpha_n^W)\right\|_F^2$, $f_n = \left\|(\alpha_n^w - M\alpha_n^X)\right\|_F^2$

The optimization algorithm for solving (4) is summarized in Algorithm 1.

3. EXPERIMENT

We evaluated our work on two public datasets, VIPeR dataset [17] and QMUL GRID dataset [18].

3.1. Datasets

The VIPeR [17] dataset contains 632 identities and each has two images captured outdoor from two views with distinct view angles. The QMUL GRID dataset [18] consists of person images captured from 8 disjoint cameras. The probe set contains 250 pedestrians, with each one having a matching image in the gallery set. Besides, there are 775 additional images in the gallery set that do not match any person in the probe set.

Table 1. Testing results on the VIPeR dataset (P=316). The CMC (%) at rank 1, 10, and 20 are listed

Method	rank=1	rank=10	rank=20
SLSSCDL (ours)	**42.77**	**85.37**	**93.76**
LSSCDL [8]	42.66	84.27	91.93
LOMO+XQDA [3]	40.00	80.51	91.08
KISSME [5]	19.60	62.20	77.00
SDALF [1]	19.87	49.37	65.73
PRDC [17]	15.66	53.86	70.09
LSSCDL[8]+smooth	42.70	84.87	92.78
LSSCDL[8]+SPL	42.73	84.96	93.42

Table 2. Testing results on the QMULGRID database (P=900). The CMC (%) at rank 1, 10, and 20 are listed

Method	rank=1	rank=10	rank=20
SLSSCDL (ours)	**22.60**	**52.40**	**63.44**
LSSCDL [8]	22.40	51.28	61.20
LOMO+XQDA [3]	16.56	41.84	52.40
PRDC [17]	15.66	53.86	70.09
RankSVM[18]	10.24	33.28	43.68
LSSCDL[8]+smooth	22.50	51.87	62.58
LSSCDL[8]+SPL	22.55	52.06	63.02

3.2. Experimental Settings

To have a fair comparison with LSSCDL [7], we apply almost the same features, classifiers, parameter settings and evaluation method as LSSCDL [7]: **Features** Like LSSCDL, we extract the LOMO [3] descriptors to represent the human appearance. We also employ [3] for dimensionality reduction, which can greatly save the time and memory. **Classifier** RBF kernel binary SVM. **Evaluation** method the cumulative matching scores (CMC) at rank rank (i). **Parameter settings** For the same parameters in the LSSCDL [9], we maintain their value unchanged for fair comparison. The remaining parameters in our experiment are λ_s and μ, we set λ_s=0.1, μ=2 for all the databases.

3.3. Result and Discussion

For the experiment, we apply all the same setting as those of LSSCDL [7], but employ SLSSCDL for training and testing. In order to analysis the effectiveness of the easiness and smoothness of our algorithm, we present two simple version of our SLSSCDL, named LSSCDL +SPL and LSSCDL +smooth, which represents integrating SPL and local similarities in to LSSCDL separately.

Learning the easiness of the feature and weight pairs To demonstrate the effectiveness of the proposed learning easiness, we compare the LSSCDL +SPL with LSSCDL [7] and SLSSCDL with LSSCDL +smooth. From the table1 and table 2 we can see that the reidentification performance can be improved by proposed scheme.

Preserving local similarities: In our framework, SLSSCDL is able to capture the local similarities in each space. It can be seen as an improved version of LSSCDL [7] in person reidentification tasks for higher efficiency. From table 1, LSSCDL [7] achieves rank-1 rate of 22.66%, rank-10 rate of 84.27%, rank-20 rate of 91.93% on the QMUL GRID dataset. In contrast, with the introduction of local smoothness regulaizer, it can improve the performance by 0.1%, 0.59%, 1.38% for the rank-1,rank-10, rank-20 rate respectively.

Combing the easiness and local similarities: We compare our SLSSCDL result with that of LSSCDL, both table 1 and table 2 indicate the effectiveness of our algorithm, especially on the QMUL GRID dataset, which improve the performance by 0.2%, 1.12%, 2.44% for the rank-1,rank-10, rank-20 rate respectively.

Convergence analysis: According to [12] the proposed optimization framework in Algorithm 1 converges. Moreover, as illustrated in Fig.2, we observe that our method achieves a stable solution within a few iterations, proving the efficiency of the proposed optimization algorithm.

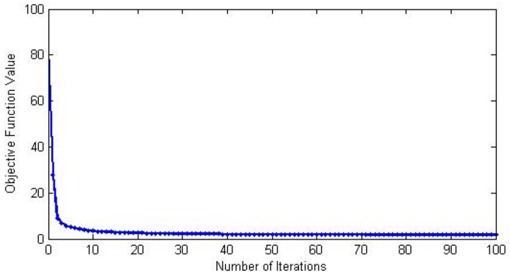

Fig. 2. The objective function value at varying number of iterations on QMUL GRID dataset.

4. CONCLUSION

In this paper, we extend LSSCDL based SampleSpecific SVM Learning for Person Re-identification to SLSSCDL. The introduced self-paced scheme is able to sort the feature and weight pairs according to both the reconstruction error in each space and the mapping coherence of the representations across space, which results in a better local minimal of the objection function. Moreover an undirected proximity graph is included to preserve the local similarities in each space. Experiments on two challenging datasets demonstrate the competitive performance of our work. Yet there are still several questions needed to be answered, such as how to adapt the abundant unlabeled data, how to use the source from other sources, etc.

5. REFERENCES

[1] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2360–2367. IEEE, 2010.

[2] Bingpeng Ma, Yu Su, and Frederic Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379–390, 2014.

[3] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identification by local maximal occurrence representation and metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2197–2206, 2015.

[4] Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*, pages 209–216. ACM, 2007.

[5] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2288–2295. IEEE, 2012.

[6] Sateesh Pedagadi, James Orwell, Sergio Velastin, and Boghos Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3318–3325, 2013.

[7] Ying Zhang, Baohua Li, Huchuan Lu, Atshushi Irie, and Xiang Ruan. Sample-specific svm learning for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1278–1287, 2016.

[8] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person re-identification by probabilistic relative distance comparison. In *Computer vision and pattern recognition (CVPR), 2011 IEEE conference on*, pages 649–656. IEEE, 2011.

[9] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, Tao Xiang, and Q Mary. Person re-identification by support vector ranking. In *BMVC*, volume 2, page 6, 2010.

[10] Wei Xu, Yijun Li, Chen Gong, and Jie Yang. Multi-view implicit transfer for person re-identification. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 1151–1155. IEEE, 2015.

[11] M Pawan Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. In *Advances in Neural Information Processing Systems*, pages 1189–1197, 2010.

[12] Deyu Meng, Qian Zhao, and Lu Jiang. What objective does self-paced learning indeed optimize? *arXiv preprint arXiv:1511.06049*, 2015.

[13] Sumit Basu and Janara Christensen. Teaching classification boundaries to humans. In *AAAI*, 2013.

[14] Miao Zheng, Jiajun Bu, Chun Chen, Can Wang, Lijun Zhang, Guang Qiu, and Deng Cai. Graph regularized sparse coding for image representation. *IEEE Transactions on Image Processing*, 20(5):1327–1336, 2011.

[15] Jingkuan Song, Yi Yang, Zi Huang, Heng Tao Shen, and Richang Hong. Multiple feature hashing for real-time large scale near-duplicate video retrieval. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 423–432. ACM, 2011.

[16] Lu Jiang, Deyu Meng, Shoou-I Yu, Zhenzhong Lan, Shiguang Shan, and Alexander Hauptmann. Self-paced learning with diversity. In *Advances in Neural Information Processing Systems*, pages 2078–2086, 2014.

[17] Douglas Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, volume 3, 2007.

[18] Chen Change Loy, Tao Xiang, and Shaogang Gong. Multi-camera activity correlation analysis. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1988–1995. IEEE, 2009.