Person Re-identification with Discriminative Dictionary Learning

Hao Sheng; Xiao Zhou, Yanwei Zheng*, Yang Liu and Da Yang
School of Computer Science and Engineering, BeiHang University, 100191 Beijing, China

ABSTRACT

Person re-identification (re-id) is important for video surveillance, and has interestingly algorithm challenges and extensive practical applications. Recently, the Sparse Representation based Classification (SRC) produced excellent results in person re-identification, in which the Dictionary Learning (DL) method is a very important part. Discriminative power of the learned dictionary determines the performance of re-identification. Previous approaches usually discriminatively train the dictionary by enforcing explicit constraints on DL. In this paper, we propose a heuristic discriminative dictionary learning method which improves the discriminative power of dictionary by transforming the representation space of dictionary in training. First, we figure out the statistical distribution of the training data and divide the data into two categories. Second, a transformation function is used to change the dictionary’s expression space. The dictionary learned by our method is proved to be effective in person re-id. Experiments on the benchmark dataset (CAVIAR4REID, i-LIDS) demonstrate that the proposed method outperforms the state-of-the-art approaches.

INTRODUCTION

Person re-identification refers to a task of associating the person between non-overlapping camera [1]. The challenging problem of person re-identification is that a person observed from different camera views with significant variations on viewpoint, pose, illumination, background clutters and occlusions. Traditional methods focused on obtaining more distinct and reliable feature representations [2][3] from people's appearances or developed machine-learning methods [4][5] that classified different persons, but it was too complicated to find features that could be applied to all the different images, just as it is currently too difficult to find a common metric model that is suitable for all scenarios. Unlike the previous methods, the Sparse Representation based Classification (SRC) models sparse coding as sparsity constrained regression problem and recently achieved the most impressive results on record for person re-id.

Sparse coding techniques are based on the fact that a nature image can be represented by a small number of atoms, which are chosen out of an over-complete dictionary. Dictionary Learning achieved impressive results in several classification and recognition problems [6] [7]. Its aim is to learn from the training images and to develop a strong representational power, and the constraints placed on it in the training process ensure that it is sparse and discriminative. Traditionally, the dictionary was obtained by enforcing explicit constraints on DL.
In this paper, we present an improved discriminative dictionary learning method which produces a more discriminative dictionary. In order to learn that dictionary, we focused on three aspects: (1) We analyze distribution of the training set and try to find the statistical curve that fits the distribution, which is proved to be subject to chi-square distribution as shown in Figure 1. (2) We introduce a piecewise function with parameters derived from the curve above which shows which part of the dictionary to stretch and which to shrink as shown in Figure 2. (1) The representation space of the dictionary is transformed in each iteration with the piecewise function. According to the statistical curve, training data are divided into two categories for transformation: the more informative data that contribute most to the dictionary's representation whereas the less informative ones make a small. We stretch the representation space that contains the more informative data and shrink the remaining space.

RELATED WORK

Existing methods for person re-id mainly focus on two aspects: appearance modeling and distance learning. Chen et al.[8] learned a similarity function to maximize the difference between the similarity scores of matched and unmatched images for a same person. Some other studies classified images into person categories with convolutional neural networks. Shi et al.[9] proposed a novel moderate positive sample mining method to train robust CNN for person re-identification, dealing with the problem of large variation and improved the learning by a metric weight constraint. Franco et al. [10] classifies images into person categories with convolutional neural networks (CNNs), and then extracts features to calculate and rank similarity of images. There were some unsupervised methods for person re-identification. Ye et al. [11] proposed a dynamic graph matching (DGM) method to estimate cross-camera labels for unsupervised re-identification.

Recently, sparse representation based classification [12] (SRC) has been successfully applied to person re-identification. Dictionary learning has been shown to provide promising results in face recognition and object classification. Sparse representation based classification (SRC) led new results in person re-id. The authors exploit the discriminative nature of sparse representation so as to perform classification. Dictionary learning is capable of discriminatively and sparsely encoding features that represent different people and the dictionary learning method is vital to sparse coding.
Dictionary learned from the training sample achieved impressive results in several classification and recognition problems [13]. Zhang and Li extended K-SVD algorithm [14] by learning an over-complete dictionary from a training dataset of nature image patches. Xu et al. [15] proposed a cloud removal method which is developed by combining multi-temporal and dictionary learning methods. The method used to recover the data contaminated by thin and thick clouds or cloud shadows. Liu et al. [16] proposed semi-supervised coupled dictionary learning to bridge the human appearance variations across cameras, and also introduced a new approach to address the person re-identification problem in cameras with non-overlapping fields of view [17]. However, the above DL methods changed the constraints to get the dictionary and treated all the features in the dictionary as identical.

To improve the discriminative power of dictionary, we propose a Discriminative Dictionary Learning method in this paper. Inspired by the statistical results of training data, we introduce piecewise function to change the dictionary’s information space in training process. Dictionary with strong presentation is learned from our method. Our experiments in bench-mark databases show that our method achieves better performance than existing dictionary learning methods.

DISCRIMINATIVE DICTIONARY LEARNING

In this section, we briefly introduce the DL based re-id methods, then present our approach to learn a discriminative dictionary and the classification scheme for re-id. We transform the dictionary in each iteration and present the dictionary visualization results as shown in Figure 2(2): background interference is reduced and the contours of the human body are more clearly. The comparison of the visualization results show that the dictionary transformation is both necessary and effective. There are two modalities for performance evaluation: the multi-versus-single (MvsS) modality and the multi-versus-multi (MvsM) modality.

Dictionary Learning Based Re-ID Methods

Dictionary learning try to obtain a dictionary $D$ and coding coefficient $X$. Let $A = [A_1; A_2; \ldots; A_n]$ be the training set, where $n$ is the number of identities.
$X = [X_1; X_2; \ldots; X_n]$ is the coefficient vector. $D$ should be able to accurately represent the person by $X$, subject to $A \approx DX$ is the sub-dictionary, $X_i$ is the coding coefficients of $A_i$ over $D$, and there is $A_i \approx DX_i = D_1X_1^i + D_2X_2^i + \ldots + D_nX_n^i$, $i = 1, 2, \ldots, n$.

Traditional dictionary learning methods focused on the constraints, and different constraints in the training process led to dictionaries with different discriminative power. One of the minimization problems researchers often use to formulate the dictionary learning process is:

$$J_{(D,x)} = \arg \min_{(D,x)} \sum_{i=1}^{n} \left( \|A_i - DX_i\|_F^2 \right)$$

$$+ \lambda_1 \|A_i - D_jx^i_j\|_F^2$$

$$+ \lambda_2 \sum_{j=1, j\neq i}^{n} \|D_jx^i_j\|_F^2$$

The learned dictionary can be used to code input person images. The coefficients will be got by solving $\hat{X} = \arg \min_x \{\|A - DX\|_2 + \lambda \|X\|_i\}$, where $\lambda$ is a scalar constant. Compute the final identity via $\text{identity}(A) = \arg \min_i \|A - DX_i\|_2$.

### Optimization of Coefficients and Dictionary

In our method, we change the representation of different characteristics in the dictionary of space to increase the represent ability of the dictionary. When updating the coefficients $X_i$, $D$ is fixed and all $X_i, i \neq j$, are fixed. Thus Eq.(1) can be reduced to:

$$J_x = \arg \min_x \sum_{i=1}^{n} \left( \|A_i - DX_i\|_F^2 \right)$$

$$+ \lambda_1 \|A_i - D_jx^i_j\|_F^2$$

$$+ \lambda_2 \sum_{j=1, j\neq i}^{n} \|D_jx^i_j\|_F^2$$

The concrete meaning of three terms in $J$ are described as: $\|A_i - DX_i\|_F^2$ express that the input signal $A_i$ can be well represented by dictionary $D$, so its value is small. $\|A_i - D_jx^i_j\|_F^2$ present that sub-set $A_i$ can be well represent by sub-dictionary $D_j$, it is expected that the images can be well represented by the images belong the same class, which implies that there are some significant coefficients in $X^i$, thus it should be small. $\sum_{j=1, j\neq i}^{n} \|D_jx^i_j\|_F^2$ means that images can't be well represented by images belong different classes. For $A_i$, $x^i_j$ should have nearly zero coefficients, thus it is small. According to the data distribution in the dictionary, the corresponding transformation
Figure 3. Performance comparison on the CAVIAR4REID dataset. MvsM when N=3 and N=5 and MvsS when N=3 and N=5.

function of different intervals was different, so the mapping was a piecewise function. So the function in Eq.(2) is deduced to:

$$J_x = \arg \min_x \sum_{i=1}^{n} \left( \| A_i - f(D)x_i \|^2 + \lambda_1 \| A_i - f(D)x_i' \|^2 + \lambda_2 \sum_{j=1, j\neq i}^{n} \| f(D_j)x_i' \|^2 \right)$$

Eq.(3) is a quadratic programming problem and it can be solved by some standard convex optimization techniques. In our model, we use the algorithm in [18].

The second important optimization problem of dictionary learning (DL) is updating $D$ with the coefficient $X$ fixed. We update $D$ class by class and when updating $D_i$, all $D_j, i \neq j$ are fixed. Eq.(1) is reduced to as follow:

$$J_{(D_i)} = \arg \min_{(D_i)} \sum_{i=1}^{n} \left( \| A_i - Dx_i \|^2 + \lambda_1 \| A_i - D_x^i \|^2 + \lambda_2 \sum_{j=1, j\neq i}^{n} \| D_jx_i' \|^2 \right)$$

In order to deduce the optimization function, we take $d_i$ as an example to calculation and derivation. $A_i - \sum_{d_u \in D} d_u X_u$, $A_i - \sum_{d_v \in D_i} d_v X_v$, and $\sum_{d_p \in D_j, d_p \neq D_i} d_p X_p$.
abbreviated to $Y_1, Y_2$ and $Y_3$, respectively, are fixed. The minimization of Eq.(4) is a single-variable optimization problem. Using the Lagrange multiplier, the optimal solution can be calculated by

$$d_i = \frac{\sum_{i=1}^{n} 2k(Y_i \beta_i^T + \lambda_i Y_i \beta_i^T - 2b_1 \beta_i^T) - \lambda_2 \sum_{i=1,j=m}^{n} 2k(Y_i \beta_i^T + b \beta_i^T)}{\left\| \sum_{i=1}^{n} 2k(Y_i \beta_i^T + \lambda_i Y_i \beta_i^T - 2b_1 \beta_i^T) - \lambda_2 \sum_{i=1,j=m}^{n} 2k(Y_i \beta_i^T + b \beta_i^T) \right\|_2}$$

(5)

Given the testing sample $A$, the sparse coding coefficients over $D$ can be get by

$$\hat{X} = \arg\min_{X} \{||y - DX||^2_2 + \gamma \||X||_1\}$$

where $\gamma$ is a constant. Serial number of the class which $A$ belongs to is the index of $D_i$ and $X_i$ that minimize $||A - D_i\hat{X}_i||_2$.

**EXPERIMENT**

This section evaluated our approach on two public available datasets, i.e. CAVIAR4REID, i-LIDS. We compared our method with some approaches including HPE[19], AHPE[20], SDALF[21], CPS[22], NSC[14], RWSC[23], MRCG[24], and reported the quantitative results in standard Cumulated Matching Characteristics (CMC) curves. Each image was processed in RGB space, resized to the same size (30*75 pixels), and represented as a vector with the dimensions $3*30*75=6750$. We randomly chose N images from individuals as the training sets (the total number of images in a training set was $N \times$ number of individuals), and the remaining images were for the testing set. In our method, we set the regularization parameters $\lambda_1=\lambda_2=1$ and $I=40$. We repeated all of the experiments 10 times and compared the average results to state-of-the-art methods.

**Performance on the CAVIAR4REID dataset**: CAVIAR4REID were captured from two cameras in a shopping center for person tracking and detection evaluations. From this Figure 3 we seen that we slightly outperform other approaches, while we significantly outperform competing methods by nearly 17% for $MvM$ ($N=3$) and by 4% for $MvS$ ($N=5$). For $MvS$, only RWSC made experiments in CAVIAR4REID, our method achieved 73.5% ($N=3$) and 80.1% ($N=5$) at rank-1.

![Figure 4. Results on i-LIDS dataset. MvsM when N=2 and MvsS when N=2.](image-url)
Performance on the iLIDS-VID dataset: iLIDS-VID was created at a crowded airport arrival hall. It consists of 600 image sequences of 300 different pedestrians that from two non-overlapping camera views. We improved on the state-of-the-art at rank-1 by nearly 5% for MvsM and by 4% for MvsS, the CMC curves are reported in Figure 4. The main complexity of the dataset arises from the very severe resolution and lighting changes between the two camera views. As can be seen, our method achieved 85.8% (MvsM) and achieved 83% (MvsS) at rank-1.

CONCLUSIONS

In this paper, we proposed a discriminative dictionary learning method for person re-identification. We conducted probability statistics of training set to divide the data into two categories, which led to the different contribution to the representation ability.

A piecewise function is introduced to transform the representation space of dictionary in training process. With the dictionary transformation, the learned dictionary enhanced the more informative part of the representation space and downgraded the less informative part. Thus the discrimination of dictionary was increased. Experimental results on three public datasets demonstrate the effectiveness of our model for person re-identification problem.

ACKNOWLEDGMENT

This study is partially supported by the National Key R&D Program of China (No.2017YFB1002000), the National Natural Science Foundation of China (No.61472019), the Macao Science and Technology Development Fund (No.138/2016/A3), the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant SKLSDE-2017ZX-09. The authors would like to thank the support from HAWKEYE Group.

REFERENCES