IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 30, 2017, accepted January 4, 2018, date of publication January 18, 2018, date of current version March 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2795020

Relative Distance Metric Leaning Based on Clustering Centralization and Projection Vectors Learning for Person Re-Identification

TONGGUANG NI[1], ZONGYUAN DING[1], FUHUA CHEN[1,2], AND HONGYUAN WANG[1]
[1]School of Information Science and Engineering, Changzhou University, Changzhou 213164, China
[2]Department of Nature Science and Mathematics, West Liberty University, West Liberty, WV 26074, USA

Corresponding author: Hongyuan Wang (e-mail: hywang@cczu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61502058 and Grant 61572085

ABSTRACT Existing projection-based person re-identification methods usually suffer from long time training, high dimension of projection matrix, and low matching rate. In addition, the intra-class instances may be much less than the inter-class instances when a training data set is built. To solve these problems, a novel relative distance metric leaning based on clustering centralization and projection vectors learning is proposed. When constructing training data set, the images of a same target person are clustering centralized with fuzzy c-means). The training data sets are built by these clusters in order to alleviate the imbalanced data problem of the training data sets. In addition, during learning projection matrix, the resulted projection vectors can be approximately orthogonal by using an iteration strategy and a conjugate gradient projection vector learning method to update training data sets. Experimental results show that the proposed algorithm has higher efficiency. The matching rate can be significantly improved, and the time of training is much shorter than most of existing algorithms of person re-identification.

INDEX TERMS Person re-identification, distance centralization, metric learning, projection vectors, conjugate gradient.

I. INTRODUCTION

Recently, surveillance systems have become ubiquitous in public places like airports, railway stations, college campuses, and office buildings [1]. There are a large number of cameras in the surveillance systems and they provide huge amounts of video data. The analysis of the computer vision obtained in a surveillance system often requires the ability to track people across multiple cameras. Therefore, person re-identification (Re-ID) has attracted more and more interests [2]–[4]. Re-ID is defined as a process of establishing correspondence between images of a person taken from different cameras. In the past five years, a large number of models have been proposed for Re-ID systems [5], [6], which can be categorized generally into two types: 1) designing discriminative, descriptive and robust visual descriptors to characterize a person's appearance [7], [8]; 2) learning suitable distance metrics that maximize the chance of a correct correspondence [9]–[11]. In this paper, we focus on the second type of person re-identification, that is, given a set of features extracted from each person image, we seek to quantify and differentiate these features by learning the optimal distance measure that is most likely to give correct matches.

Many metric learning algorithms have been proposed to act with the distances or similarity functions of person re-identification features. For example, Pedagadi et al. applied the local fisher discriminate analysis (LFDA) [12] to solve person re-identification problems. The authors in [13] introduced the KISSME method from equivalence constraints based on a statistical inference perspective. Dikmen et al. proposed a metric learning framework to obtain a robust Mahalanobis metric for large margin nearest neighbor classification with rejection (LMNN) [14]. Davis *et al.* [15] presented an information-theoretic approach to learn a Mahalanobis distance function. Zheng et al. proposed the relative distance comparison (RDC) approach to maximize the likelihood of a pair of true matches which having a relatively smaller distance than that of a wrongly matched pair in a soft discriminate manner [16]. Chen *et al.* [17] formulated an asymmetric distance model for learning camera-specific projections to transform the unmatched features of each view

to a common space. Chen *et al.* [18] presented a relevance metric learning method with listwise constraints (RMLLCs) by adopting listwise similarities using predefined similarity lists.

However, the existing Re-ID methods discussed above all have two shortcomings.

1) As we all know, for the person re-identification datasets, there are usually much more inter-class person image pairs than intra-class ones. Imbalanced datasets would drastically affect the modeling of machine learning methods.

2) The training set for learning the model consists of images of matched people across different camera views. In order to capture the large intra and inter variations, this represents a large scale learning problem that challenges existing machine learning algorithms.

To address the problems of the existing Re-ID methods, a novel method, called relative distance metric leaning based on clustering centralization and projection vectors learning for person re-identification (RDML-CCPVL) is proposed. First, there are usually much more inter-class instances than intra-class instances when traditional metric learning methods collect training dataset. Constructing counter examples of each instance needs to compute the distances with all the other instances, so the training time is greatly increased. Using FCM [19], the number of counterexamples of each instance is decreased by divided into clusters and the important structural information is retained, so the overfitting problem caused by class imbalance is relieved.

Second, traditional matrix projection learning methods usually have greater storage and computing complexity. In this work, we decomposed the projection matrices into low rank ones by eigenvalue decomposition for projection matrices. According to our iterative optimization method, updating the distance vectors of instance features only need to learn a new projection vector using the updated training dataset each time and stop when achieving a good enough accuracy. The conjugate gradient method is used to learn the projection vector, which only needs to compute the initial gradient one time. For the quadratic function, the conjugate gradient method can converge to the target precision soon due to quadratic termination. Our method can effectively reduce the computational complexity and storage. In addition, our algorithm can approximately ensure to keep the orthogonal characteristics of the vectors after eigenvalue decomposition.

II. RELATIVE DISTANCE METRIC LEANING BASED ON CLUSTERING CENTRALIZATION AND PROJECTION VECTORS LEARNING

A. LEARNING FUNCTION BASED ON CLUSTERING CENTRALIZATION

The person re-identification problem can be casted into a distance comparison problem [16]. Suppose we have a set of m training pedestrian images $D^k = [\boldsymbol{X}, \boldsymbol{Y}]_1^m$ with a feature dataset $\boldsymbol{X} = \{\boldsymbol{x}_i, i = 1, \ldots, m\}, \boldsymbol{x}_i \in R^d$, where d is the dimension of the features and $y = \{y_i, i = 1, \ldots, m\}$ is the input label dataset. For an instance $\boldsymbol{x}_a$ of person A, we want to find another instance $\boldsymbol{x}_b$ of person A captured elsewhere in space and time by learning the distance $dis(\boldsymbol{x}_a, \boldsymbol{x}_b)$ of these two instances. As we all know, the distances of intra-class instances should be smaller in general compared with the distances of inter-class instances, so that $dis(\boldsymbol{x}_a, \boldsymbol{x}_b) < dis(\boldsymbol{x}_a, \boldsymbol{x}_c)$, where $\boldsymbol{x}_c$ is an instance of any other pedestrian except A. Here we construct a pairwise dataset $\mathrm{S} = \{\mathrm{S}_t = (\boldsymbol{d}_t^{pos}, \boldsymbol{d}_t^{neg})\}_{t=1}^m$ to describe the distances of the instances, where $\boldsymbol{d}_t^{pos}$ is the distance of $\boldsymbol{x}_t$ with intra-class instances and $\boldsymbol{d}_t^{neg}$ is the distance of $\boldsymbol{x}_t$ with inter-class instances.

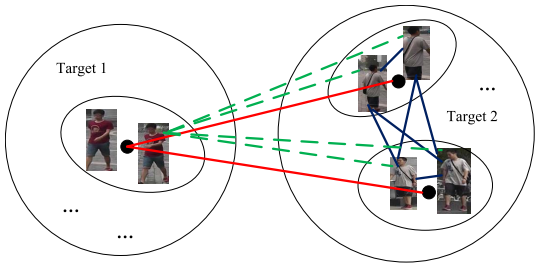


FIGURE 1. The principle of the clustering centralization.

Obviously, the number of inter-class instances is much larger than the number of intra-class instances. So the pairwise dataset $\mathrm{S} = \{\mathrm{S}_t = (\boldsymbol{d}_t^{pos}, \boldsymbol{d}_t^{neg})\}_{t=1}^m$ will be a class imbalanced dataset. Training with such a dataset will lead to over fitting of the inter-class instances and under fitting of the intra-class instances, which will decrease the performance of the learning algorithm. Here we choose FCM [19] to alleviate the class imbalanced data problem. The principle of our clustering centralization is shown in Fig.1. From Fig.1, the traditional training datasets constructing methods [16] (shown with green lines) include much more inter-class instances, while our method (shown with red lines) using clustering centralization can effectively alleviate the class imbalanced problem of the existing Re-ID algorithms. The other advantage of our training clustering centralization method is the number of clusters can adjust for different Re-ID datasets to get the optimal performance, which will be discussed in section 4.3 lately.

After performing clustering centralization, we obtain a set of pairs of distances, $\mathrm{S} = \{\mathrm{S}_t = (\boldsymbol{d}_t^{pos}, \bar{\boldsymbol{d}}_t^{neg})\}_{t=1}^m$, where $\bar{\boldsymbol{d}}_t^{neg}$ is the distance of $\boldsymbol{x}_t$ with the centers of clusters of its counterexamples. In order to maximize the inter-class distance and minimize the intra-class distance at the same time, we can formulate it into a minimization problem as following:

$$dis(\boldsymbol{d}_t^{pos}, \bar{\boldsymbol{d}}_t^{neg}) = g(\boldsymbol{d}_t^{pos}) - g(\bar{\boldsymbol{d}}_t^{neg}) \tag{1}$$

where $g(\cdot)$ is a distance function. The function in (1) is unbounded, so it cannot guarantee convergence during iteration. Here, we transform it into a continuous sigmoid function as following:

$$dis(\boldsymbol{d}_t^{pos}, \bar{\boldsymbol{d}}_t^{neg}) = (1 + \exp(g(\boldsymbol{d}_t^{pos}) - g(\bar{\boldsymbol{d}}_t^{neg}))^{-1} \tag{2}$$

Considering the convenience of computation, Equation (1) is then transformed to a logistic form as following;

$$\begin{aligned} f &= -\log(\prod_{S_t} dis(\boldsymbol{d}_t^{pos}, \bar{\boldsymbol{d}}_t^{neg})) \\ &= \sum_{S_t} \log(1+\exp(g(\boldsymbol{d}_t^{pos}) - g(\bar{\boldsymbol{d}}_t^{neg}))) \end{aligned} \tag{3}$$

We can see minimizing (1) is equivalent to maximizing (2) and maximizing (2) is equivalent to minimizing (3). Because the Mahalanobis matrix of the Mahalanobis distance function has good projective property and learning property, here we choose the Mahalanobis distance function as the distance function g:

$$g(\boldsymbol{d}_t^{pos}) = (\boldsymbol{d}_t^{pos})^T \boldsymbol{M} (\boldsymbol{d}_t^{pos}) \tag{4}$$

where $\boldsymbol{M}$ is a semi-definite matrix. Our goal becomes to learn $\boldsymbol{M}$ in (4) by minimizing the functional defined in (3). By performing eigenvalue decomposition on $\boldsymbol{M}$, we can find $\boldsymbol{M} = \boldsymbol{P}\boldsymbol{P}^T$, $\boldsymbol{P}$ is a matrix of column orthogonal vectors. The number of orthogonal bases may be smaller than the rank of matrix $\boldsymbol{M}$. Therefore, $\boldsymbol{P} \in \mathrm{R}^{n*d'}$ can be regard as a dimension reduction matrix, where d' is the number of orthogonal basis after dimension reduction. Equation (4) can be transformed into following:

$$\begin{aligned} g(\boldsymbol{d}_t^{pos}) &= (\boldsymbol{d}_t^{pos})^T \boldsymbol{M} (\boldsymbol{d}_t^{pos}) \\ &= (\boldsymbol{d}_t^{pos})^T \boldsymbol{P}\boldsymbol{P}^T (\boldsymbol{d}_t^{pos}) = \left\| \boldsymbol{P}^T \boldsymbol{d}_t^{pos} \right\|^2 \end{aligned} \tag{5}$$

In addition, for a small dataset, the function shown in (3) may be an overfitting learning problem. In order to alleviate the risk of over fitting and ensure the sparsity of the projection matrix, we introduce $r \left\| \boldsymbol{P} \right\|^2$ as a regularization term, where r is the regularization factor. The distance function can be formulated as:

$$f = \sum_{S_t} \log(1+\exp(\left\| \boldsymbol{P}^T d_t^{pos} \right\|^2 - \left\| \boldsymbol{P}^T \bar{d}_t^{neg} \right\|^2)) + r \left\| \boldsymbol{P} \right\|_2^2 \tag{6}$$

B. AN ITERATIVE OPTIMIZATION ALGORITHM FOR PROJECTION VECTOR LEARNING

In this paper, we choose a similar iterative optimization method of [16] to learn an optimal $\boldsymbol{P}$. Starting from an empty matrix, a new estimated column $\boldsymbol{p}_l$ will added to $\boldsymbol{P}$ after lth iteration. Each iteration consists of two steps as follows:

Step 1: Assume that after l-1 iterations a set of orthogonal vectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_{l-1}$ have been learned, to the next vector $\boldsymbol{p}_l$, let

$$\boldsymbol{d}_t^{s,l} = \boldsymbol{d}_t^{s,l-1} - \frac{\boldsymbol{p}_{l-1}\boldsymbol{p}_{l-1}^T}{\left\| \boldsymbol{p}_{l-1} + u \right\|^2} \boldsymbol{d}_t^{s,l-1} \tag{7}$$

where $l > 1$, u is small perturbation number. We defined $\boldsymbol{d}_t^{s,0} = \boldsymbol{d}_t^s$, $s \in \{pos, neg\}$, $t \in 1, \ldots, |\mathrm{S}|$.

Step 2: After obtain $d_t^{s,l}$ from (7), let $\mathrm{S}^l = \{\mathrm{S}_t^l = (\boldsymbol{d}_t^{pos,l}, \bar{\boldsymbol{d}}_t^{neg,l})\}$. Then, we use conjugate gradient function $f(\boldsymbol{p}_l^k)$ to learn projection vectors of (8).

$$f(\boldsymbol{p}_l^k) = \sum_{d_t \in \mathrm{S}^l} \log(1+\exp(\left\| (\boldsymbol{p}_l^k)^T \boldsymbol{d}_t^{pos} \right\|^2 - \left\| (\boldsymbol{p}_l^k)^T \bar{\boldsymbol{d}}_t^{neg} \right\|^2)) + r \left\| \boldsymbol{p}_l^k \right\|^2 \tag{8}$$

The gradient of $f(\boldsymbol{p}_l^k)$ is computed by

$$\begin{aligned} g_l = \frac{\partial f}{\partial \boldsymbol{p}_l} = \sum_{d_t \in \mathrm{S}^l} & \frac{2\exp(\left\| (\boldsymbol{p}_l^k)^T \boldsymbol{d}_t^{pos} \right\|^2 - \left\| (\boldsymbol{p}_l^k)^T \bar{\boldsymbol{d}}_t^{neg} \right\|^2)}{1+\exp(\left\| (\boldsymbol{p}_l^k)^T \boldsymbol{d}_t^{pos} \right\|^2 - \left\| (\boldsymbol{p}_l^k)^T \bar{\boldsymbol{d}}_t^{neg} \right\|^2)} \\ & \times (\boldsymbol{d}_t^{pos} \boldsymbol{d}_t^{pos^T} - \bar{\boldsymbol{d}}_t^{neg} \bar{\boldsymbol{d}}_t^{neg^T}) \boldsymbol{p}_l^k + 2r\boldsymbol{p}_l^k \end{aligned} \tag{9}$$

The optimal projection vector after kth iteration is defined as following:

$$\boldsymbol{p}_l^{k+1} = \boldsymbol{p}_l^k + \alpha_k q_k, \tag{10}$$

where α_k is computed by $f(\boldsymbol{p}_l^k + \alpha q_k)$ using one-dimensional accurate search, $\boldsymbol{q}_k$ is the search direction of projection vector after kth iteration, and the conjugate direction is computed by PRP equation as following

$$q_k = -g_k + \beta_{k-1} q_{k-1}, \beta_{k-1} = \begin{cases} \dfrac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}, & k > 1 \\ 0, & k = 1 \end{cases} \tag{11}$$

If $|f(p_l^k) - f(p_l^{k-1})| < \varepsilon_g$, the iteration is terminated.

The initial value of p_l is formulated as following:

$$\boldsymbol{p}_l^0 = \frac{1}{\left\| \mathrm{S}_{pos}^l \right\|} \sum_{\mathrm{S}_{pos}^l} \boldsymbol{d}_i^{pos} - \frac{1}{\left\| \mathrm{S}_{neg}^l \right\|} \sum_{\mathrm{S}_{neg}^l} \bar{\boldsymbol{d}}_i^{neg} \tag{12}$$

According to (9) and (11), we can see $\boldsymbol{p}_l \in span\{\boldsymbol{d}_i^{s,l}\}$, where $span\{\boldsymbol{d}_i^{s,l}\}$ is a range space of $\{\boldsymbol{d}_i^{pos,l}\} \cup \{\bar{\boldsymbol{d}}_i^{neg,l}\}$, $s \in \{pos, neg\}$, $i \in 1, \ldots, |\mathrm{S}|$. According to (7), we know that $\boldsymbol{p}_j^T \boldsymbol{d}_i^{s,j+1} \approx 0$ where $j = 1, \ldots, l-1$ and $\boldsymbol{p}_l \in span\{\boldsymbol{d}_i^{s,l}\}$, $span\{\boldsymbol{d}_i^{s,l}\} \subseteq span\{\boldsymbol{d}_i^{s,l-1}\} \subseteq \ldots \subseteq span\{\boldsymbol{d}_i^{s,0}\}$, so $\boldsymbol{p}_l$ and $\boldsymbol{p}_j$, $j = 1, \ldots, l-1$ are approximately orthogonal.

Different from the iterative optimization algorithm of [16], a smaller perturbation term u is added to (7), which makes each projection space preserving a relation with each other and is more suitable to the real-world learning problem.

III. LEARNING ALGORITHM FOR RDML-CCPVL

Based on the above iterations, the learning algorithm of the proposed RDML-CCPVL is presented in Algorithm 1.

IV. EXPERIMENTS AND ANALYSIS

A. DATASETS AND EXPERIMENTAL SETTING

Six popular datasets are selected for our experiments: VIPeR [20], CUHK01 [21], 3DPeS [22], CAVIAR4REID [23], Town Centre [24], and Market-1501 [25]. The VIPeR dataset consists of 632 pedestrian image pairs taken from two camera views (Fig. 2(a)).

Algorithm 1 Learning Algorithm for RDML-CCPVL

Input: $X = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$, u, ε_o, ε_g
begin:
build $\mathrm{S} = \{\mathrm{S}_t = (\boldsymbol{d}_t^{pos}, \bar{\boldsymbol{d}}_t^{neg})\}_{t=1}^{m}$ with X using clustering centralization;
$\boldsymbol{P} = [\,]$, $\boldsymbol{p}_0 = 0$, $l = 0$;
while true
$l = l + 1$;
update S^l with (7);
compute $\boldsymbol{p}_l^0$ using (12);
$k = 0$;
compute $f(\boldsymbol{p}_l^k)$ using (8);
while true
$k = k + 1$;
compute g_l using (9);
compute q_k using equation 10;
compute β_{k-1} using equation 11;
update $\boldsymbol{p}_l^k$ with q_k;
compute $f(p_l^k)$ using equation 7;
if $|f(\boldsymbol{p}_l^k) - f(\boldsymbol{p}_l^{k-1})| \le \varepsilon_g$
break;
end if
end while
$\boldsymbol{P} = [\boldsymbol{P}, \boldsymbol{p}_l]$
compute f^l using (6);
if $\left|f^l - f^{l-1}\right| \le \varepsilon_o$
break;
end if
end while
end
Output: $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_l]$

FIGURE 2. Image samples of the five datasets. Images in the same column are from the same person across two views. (a)VIPeR; (b)CUHK01; (c)3DPeS; (d)CAVIAR4REID; (e)Town Centre; (f)Market-1501.

TABLE 1. Comparisons of performance on dataset VIPeR with c = 2 (%).

Methods	Rank 1	Rank 10	Rank 20	Rank 30	Training time(s)
RDML-CCPVL	13.33	**60.00**	**83.33**	**91.67**	**77.69**
CVDCA	**16.68**	54.11	67.73	87.5	187.56
RMLLC	15.22	47.37	59.52	85.33	502.63
LFDA	11.45	40.98	56.03	82.77	624.37
PRDC	10.36	37.62	54.14	78.35	465.26
KISSME	7.21	30.68	44.46	75.24	738.48
ITML	9.50	36.74	54.66	80.00	985.67
LMNN	8.46	34.58	50.25	74.98	872.53

TABLE 2. Comparisons of performance on dataset CUHK01 with c = 2(%).

Methods	Rank 1	Rank 10	Rank 20	Rank 30	Training time(s)
RDML-CCPVL	**18.52**	**57.14**	**79.89**	**86.67**	**98.73**
CVDCA	17.77	53.26	68.72	84.46	225.23
RMLLC	16.02	54.85	73.33	83.87	324.75
LFDA	12.47	51.15	66.33	81.72	486.8
PRDC	14.58	52.60	68.50	82.67	502.43
KISSME	10.65	48.78	62.45	79.63	798.57
ITML	9.24	45.05	61.36	75.52	1095.4
LMNN	10.46	47.76	58.45	72.00	924.06

TABLE 3. Comparisons of performance on dataset 3DpeS with c = 10 (%).

Methods	Rank 1	Rank 10	Rank 20	Rank 30	Training time(s)
RDML-CCPVL	**24.58**	**54.19**	**67.6**	**82.12**	**203.86**
CVDCA	22.54	50.44	64.71	77.64	1072.54
RMLLC	21.04	51.25	63.45	75.73	1521.82
LFDA	12.24	48.75	59.07	71.98	1795.15
PRDC	17.33	50.04	61.72	73.66	1302.25
KISSME	7.12	29.02	46.36	58.79	2577.06
ITML	10.65	31.44	54.47	61.73	4686.34
LMNN	8.42	26.7	50.09	57.93	2410.95

TABLE 4. Comparisons of performance on dataset CAVIAR4REID with c = 15(%).

Methods	Rank 1	Rank 10	Rank 20	Rank 30	Training time(s)
RDML-CCPVL	**13.19**	**56.88**	**73.44**	**89.69**	**207.83**
CVDCA	12.36	54.3	70.05	84.81	1104.71
RMLLC	10.78	53.65	71.26	88.52	1589.13
LFDA	8.67	44.29	65.15	80.38	1645.65
PRDC	9.54	54.12	70.97	82.46	1147.87
KISSME	5.75	36.12	58.98	75.18	2612.52
ITML	4.15	29.71	60.08	77.17	4963.28
LMNN	2.18	26.56	62.17	59.6	2553.46

The CUHK01 dataset contains 971 individuals also captured from two camera views (Fig. 2(b)). The 3DPeS dataset is collected by 8 non-overlapped outdoor cameras (Fig. 2(c)). The CAVIAR4REID dataset is extracted from a multi-target tracking dataset CAVIAR, which is collected in a shopping mall by two surveillance cameras with overlapped view field. Among 72 identities, 50 of them have images from two camera views and the rest 22 only from one camera (Fig. 2(d)). The Town Center dataset is a 5 min video with 7500 frames annotated, which is divided into 6500 images for training and 1000 images for testing data for pedestrian detection (Fig. 2(e)). The Market-1501 dataset is collected in front of a supermarket in Tsinghua University. A total of six cameras are used, including 5 high-resolution cameras, and one low-resolution camera. Overlap exists among different cameras. Overall, this dataset contains 32668 images of 1501persons (Fig. 2(f)).

TABLE 5. Comparisons of performance on dataset Town Centre with c = 2 (%).

Methods	Rank 1	Rank 10	Rank 20	Rank 30	Training time(s)
RDML-CCPVL	**62.22**	**86.36**	**92.11**	**95.33**	**174.41**
CVDCA	57.09	83.45	90.74	92.1	1076.05
RMLLC	55.97	80.84	88.86	93.04	1516.84
LFDA	46.89	77.82	80.92	89.15	1748.15
PRDC	52.46	81.15	86.81	90.43	1489.94
KISSME	38.47	68.17	75.52	84.63	2933.23
ITML	25.58	46.25	59.76	71.04	5364.69
LMNN	22.92	39.24	46.82	64.8	2986.14

TABLE 6. Comparisons of performance on dataset Market-1501 with c = 5(%).

Methods	Rank 1	Rank 10	Rank 20	Rank 30	Training time(s)
RDML-CCPVL	**26.95**	**78.49**	**89.94**	**94.86**	**1525.62**
CVDCA	24.02	75.4	84.82	90.54	2588.65
RMLLC	23.93	72.86	81.15	85.9	3134.43
LFDA	18.59	62.04	75.17	80.06	4912.77
PRDC	20.22	68.4	79.27	84.64	5974.45
KISSME	15.52	59.87	70.93	75.48	6124.38
ITML	12.07	52.43	68.55	70.32	7059.74
LMNN	11.25	45.89	62.24	69.58	8208.03

In our experiments, we randomly select all images of 200,70,134,30,100,1000 people classes from the VIPeR, CUHK01, 3DPeS, CAVIAR4REID, Town Centre and Market-1501 datasets respectively, to set up the training set, and the rest of the people classes were used for training. Different numbers of people classes are used to evaluate the matching performance of models learned with different amounts of training data. This procedure was repeated 10 times. During the training, a pair of images of each person formed a relevant pair, and one image of him/her and one of another person in the training set formed a related irrelevant pair, and together they formed the pairwise set S defined in Section 2. For each image, we use six type of features descriptor, such as RGB, YCbCr, HSV, Lab, YIQ and Gabor [26]. Then we use PCA to compress them into 2688-dimensional feature vectors.

We compare the proposed RDML-CCPVL with seven existing person re-identification works: ITML [15], LMNN [14], KISSME [13], PRDC [16], LFDA [12], CVDCA [17], RMLLC [18]. The performance of all the methods is evaluated in terms of cumulative matching characteristic (CMC), which is a standard measurement for Re-ID [16]. The CMC curve represents the probability of finding the correct match over the top r in the gallery image ranking, with r varying from 1 to 30. In order to evaluate the efficiency of our algorithm, in this paper, the comparisons of the training time between our algorithm with other existing algorithms are also shown. Since the number c of the clustering centers is an important parameter for our algorithm, Normalized Discounted Cumulative Gain(NDCG) [27] is choose to evaluate the performance of RDML-CCPVL with varying number of c. We run all the benchmarking algorithms with MATLAB 7 on a 1.90GHz machine with 8G RAM.

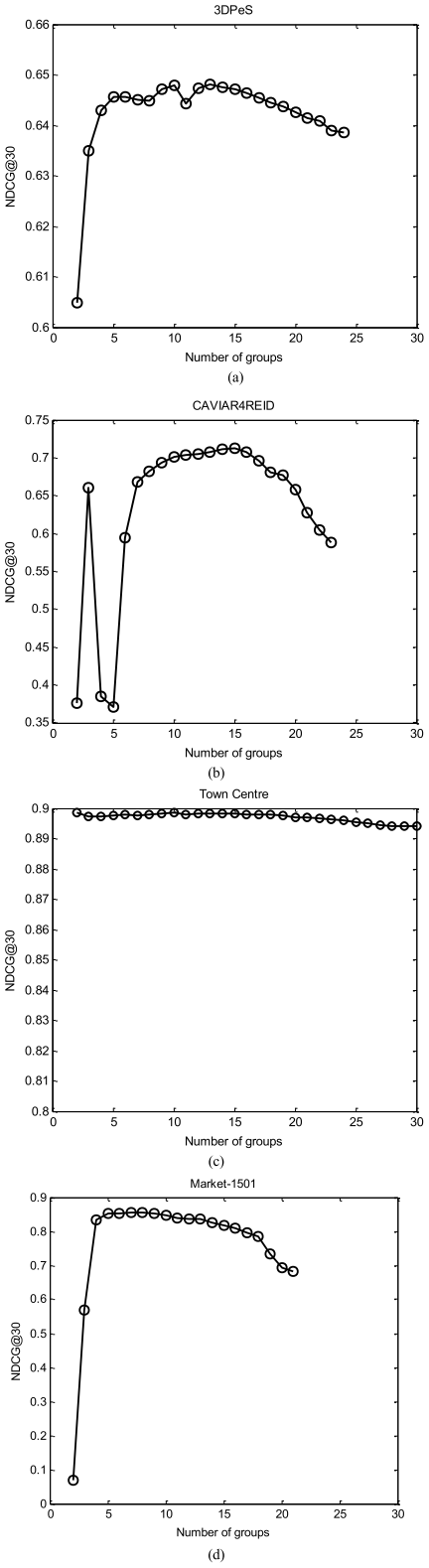


FIGURE 3. Comparison of the performance with varying number of the clusters c. (a)3DPeS; (b)CAVIAR4REID; (c)TownCentre; (d)Market-1501.

B. COMPARISONS OF PERFORMANCE ON DEFFERENT RE-ID DATASETS

Comparisons of the CMC and training time on different Re-ID datasets are shown in Table 1-Table 6 with optimal value of c, respectively. Since clustering centralization is used to solve the imbalanced data problem of the training datasets, it can be seen that the performance of our algorithm is considerably superior to other algorithms on almost all datasets. The CMC of RDML-CCPVL is at least 3.2%-4.5% higher than other algorithms at rank 30 on all datasets. The other advantage of clustering centralization is the reduction of the training datasets, we can see the training time of RDML-CCPVL is much less than other algorithms on all datasets.

C. EXPERIMETNS WITH VARY NUMBER OF CLUSTERS

In this section, we report the change tendency of the performance of RDML-CCPVL by running them on the different datasets with varying number of clusters c. Since there are very few images of the same person in VIPeR and CUHK01, we only show the figures of the other four datasets 3DPeS, CAVIAR4REID, Town Centre and Market-1501. From the Fig.3 we can see that the proposed RDML-CCPVL is considerably sensitive to c for all the datasets except Town Centre. The performance on 3DPeS, CAVIAR4REID and Market-1501 will become better as c increases because more clusters mean more information of the datasets. But when c becomes too large, the curves in Fig.3 start to go down because we find that c has an optimal range of value for different datasets and too many clusters also can not generate useful distribution information. For Town Centre, which include only images of Video sequences, so how many clusters of intra-class instances of the same person has little difference. When $c = 2$, the proposed algorithm achieves the optimal performance on Town Centre.

V. CONCLUSION

In this work, we proposed a relative distance metric leaning algorithm based on clustering centralization and projection vectors learning for person re-identification problem. We have shown that cluster centralization can improve the performance and efficiency in person re-identification and reduce both the computational complexity and the storage space. In addition, the conjugate gradient method is used in the projection vector learning. The proposed approach shows a significant improvement over other existing algorithms.

REFERENCES

[1] P. Agrawal and P. Narayanan, "Person de-identification in videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 299–310, Mar. 2011.

[2] X. Cai, C. Wang, B. Xiao, X. Chen, and J. Zhou, "Deep nonlinear metric learning with independent subspace analysis for face verification," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 749–752.

[3] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen, "Fusing robust face region descriptors via multiple metric learning for face recognition in the wild," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3554–3561.

[4] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[5] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2707–2714.

[6] J. Hu, J. Lu, and Y. P. Tan, "Discriminative deep metric learning for face verification in the wild," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1875–1882.

[7] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2197–2206.

[8] L. An, M. Kafai, S. Yang, and B. Bhanu, "Person re-identification with reference descriptor," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 776–787, Apr. 2016.

[9] L. An, S. Yang, and B. Bhanu, "Person re-identification by robust canonical correlation analysis," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1103–1107, Aug. 2015.

[10] C. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," in *Proc. 20th IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 3567–3571.

[11] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznai, and H. Bischof, "Mahalanobis distance learning for person re-identification," in *Person Re-Identification*. London, U.K.: Springer, 2014, pp. 247–267.

[12] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3318–3325.

[13] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2288–2295.

[14] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2010, pp. 501–512.

[15] J. V. Davis, B. Kulis, and P. Jain, "Information theoretic metric learning," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 209–216.

[16] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.

[17] Y. C. Chen, W. S. Zheng, and J. H. Lai, "An asymmetric distance model for cross-view feature mapping in person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 8, pp. 1661–1675, Aug. 2016.

[18] J. X. Chen, Z. X. Zhang, and Y. L. Wang, "Relevance metric learning for person reidentification by exploiting listwise similarities," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4741–4755, Dec. 2015.

[19] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer, 1981.

[20] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 262–275.

[21] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 31–44.

[22] D. Baltieri, R. Vezzani, and R. Cucchiara, "3DPeS: 3D people dataset for surveillance and forensics," in *Proc. Joint ACM Workshop Human Gesture Behav. Understanding*, 2011, pp. 59–64.

[23] D. S. Cheng, M. Cristani, and M. Stoppa, "Custom pictorial structures for re-identification," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 1–11.

[24] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 3457–3464.

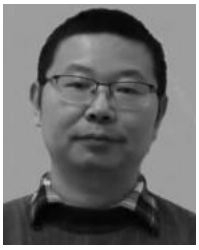
[25] L. Zheng, L. Shen, and L. Tian, "Scalable person re-identification: A benchmark. Computer vision," in *Proc. 14th IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 1116–1124.

[26] Y. Chen, W. Zheng, and J. Lai, "Mirror representation for modeling view-specific transform in person re-identification," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 3402–3408.

[27] Y. Wang, L. Wang, Y. Li, D. He, W. Chen, and T. Y. Liu, "A theoretical analysis of NDCG ranking measures," in *Proc. 26th Annu. Conf. Learn. Theory (COLT)*, 2013, pp. 1–30.

TONGGUANG NI received the Ph.D. from Jiangnan University in 2015. He is currently a Lecturer with the School of Information Science and Engineering, Changzhou University, Changzhou, China. His current research interests include pattern recognition, intelligent computation, and their application.

ZONGYUAN DING was born in Huai'an, China, in 1991. He received the B.S. degree in information and computing science from Huaiyin Normal University in 2015. He is currently pursuing the master's degree with Changzhou University. His main research interests include image processing and pattern recognition.



FUHUA CHEN received the Ph.D. degree in computer science from the Nanjing University of Science & Technology and the Ph.D. degree in applied mathematics from the University of Florida. He is currently an Assistant Professor at West Liberty University. His research interest is in variation image segmentation, mathematical modeling, and person re-identification.

HONGYUAN WANG received the Ph.D. degree in computer science from the Nanjing University of Science & Technology. He is currently a Professor with Changzhou University. His general research interests include pattern recognition and intelligence system. His current interest is in pedestrian trajectory discovery in intelligent video surveillance.