A Review for Person Re-identification

Weirao Wang

Department of Computer Science and Technology, Donghua University, 201620 Shanghai, China

Keywords: Person Re-Identification, semi-supervised learning, Deep Learning.

Abstract: Person re-identification aims to judge whether the person images obtained under different cameras correspond to the same person. It can be regarded as a subproblem of image retrieval and has broad application prospects in the fields of intelligent video surveillance, security, and criminal investigation. Person re-identification has become a challenging task in the field of computer vision and attracted more researchers due to the unstable image features affected by low-image resolution, inconsistent shooting angles, poor lighting conditions, continuous changes in background and poses, etc. Existing researches on person re-identification are mainly based on the person images dataset proposes by various institutions, and study learning feature models and measurement methods. Early researches focus on the problem of misalignment and the feature instability caused by the illumination change, trying to build a more stable feature model with stronger discriminative ability. With the deepening of research, more person re-identification datasets expose, and the sample size is constantly expanding. In recent years, with the great success of deep learning algorithm in the field of computer vision, this technology has been introduced into the research of person re-recognition algorithm, which has promoted the improvement of accuracy of person re-identification, bringing new opportunities to the development of person re-identification technology. This paper summarizes the development history, research status and typical algorithms of person re-identification. First, the basic research framework of person re-identification is explained. Then, the research results of the two key technologies (feature expression and similarity metric) of person re-identification are summarized. This paper mainly introduces and analyzes the person re-identification technology based on deep learning algorithm theory. Finally, the representative public datasets in person re-identification are introduced, and compares the performance of the algorithm on the VIPeR dataset. In this paper, based on the JSTL deep feature model, the semi-supervised method is introduced, and the algorithm is improved and analyzed.

1. Introduction

Person re-identification is widely considered as a subproblem of image retrieval. It uses computer vision technology to judge whether the person images obtained from different cameras correspond to the same person in non-overlapping monitoring network. Person re-identification technology can make up for the visual limitations of fixed cameras, and can be combined with pedestrian detection and human tracking technology to be applied to intelligent video monitoring, security and criminal investigation and other fields. Early researches on person re-identification mainly focus on how to design a better feature representation model or learn a similarity metric model with strong discriminant ability and robustness before the emergence of deep learning. In recent years, with the great success of deep learning algorithm in various computer vision tasks, deep learning method has been successfully introduced into the field of person re-identification [1-3].

Different from traditional methods, deep learning algorithms have stronger feature expression ability. Feature expression model and similarity measurement model with stronger adaptability are studied simultaneously through end-to-end learning method. At first, deep learning methods introduced into this task are simple and have poor accuracy, which is lower than most measurement learning algorithms in the same period. With the further research, some researchers continue to improve the deep network framework according to the difficulties of person re-identification.

According to the difference of loss function, it can be divided into representation learning and metric learning based method. During this period, due to the instability of global features, researchers build more complex network and extract local features on patch-based learning. At the same time, this task also develops from one-shot recognition to image sequence recognition method research. In recent years, with the success of the Generative Adversarial Nets (GAN) model in other fields, researchers carry out researches on person re-identification method based on GAN network, mainly through the GAN network model to enrich the samples and alleviate the impact of small samples on the generalization of the person re-identification model [1]. The existing person re-identification technology mainly focuses on supervised learning model, whereas transfer learning, semi-supervised learning and unsupervised learning also have important significance for the research of person re-identification algorithm. In this paper, the performance of person re-identification algorithm in recent years is summarized, and the semi-supervised person re-identification method is further studied.

2. Appearance Feature Representation Model for Person Re-identification

Due to the low resolution of person images, it is impossible to authenticate person's identity through detailed biometric (face, iris), so the color and texture information of clothing becomes the key to person image recognition. The extraction of appearance features can be divided into three categories: manual feature, mid-level feature and semantic feature.

2.1 Manual Feature

Among all the image features of person re-identification, the most stable information is color information because the color of the target's clothing is the most prominent feature in images. Common color feature expression models include color histogram and color moment. Barenzena et al.[2] proposed a Symmetry-Driven Accumulation of Local Features (SDALF) for feature representation which used symmetry of body structure to design feature model, and used color histogram to extract color features of person images. Moreover, this method combined weighted color histogram, the most stable region color and highly repetitive structure color block, these three color information features, to build a more robust color feature expression model.

However, color information is easily affected by external illumination in outdoor scenes. In addition, considering similarity of clothing color among persons (especially men) in the same season, researchers extract texture information of person's outline and clothing pattern by adding texture features. The apparent feature representation can be obtained through the combination of two information features [3]. Although many algorithms have been designed for the feature expression of person images, features change greatly due to viewpoint changes and person's misalignment. In this regard, some researchers proposed methods that divide images into patches according to certain rules, respectively do similarity measurement and then concatenate measurement results of each patch [2,4], or simply concatenate the features of each patch into feature vector for similarity measurement [3,5,6,7]. However, this method is relatively rough in processing, and the body structure in the image patch has a poor corresponding effect, so the improvement of the feature model is limited. However, this processing of manual features will make resulting feature dimension too high. In this respect, some researchers proposed methods of feature representation based on covariance [8,9,10]. Matsukawa et al. [8] proposed Hybrid Spatiogram and Covariance Descriptor (HSCD) for feature extraction, which combined the color histogram accumulated by patches in different areas with the covariance matrix of color and texture information.

2.2 Mid-level Feature

Mid-level feature is a method to extract image feature by filter. This kind of feature model learns filter operators with expressive ability through training data. Different filter operators have responses to different geometric structures and color information. For mid-level feature extraction of person images in person re-identification, saliency detection methods were introduced to extract regions

critical for target identity authentication which is called salient regions, and then used apparent features of salient regions to improve the discriminative ability and robustness of person image recognition. Zhao et al. [9] proposed a person re-identification method based on the saliency learing and saliency distribution, used K-nearest neighbor and SVM algorithm to score the saliency of image patches, and then used the matching between patches to measure similarity. In addition, Zhao et al. [11] studied the mid-level feature extraction method based on filter operator, and designed a mid-level feature filter operator. The filter operator in this method has good ability to distinguish different color and texture patches. Meanwhile, a simple and effective cross-view training strategy was proposed in this model, so that the filter has a certain invariant to viewpoint changes and strong discriminative ability.

Dictionary learning [12,13,14] is also an effective mid-level feature expression method. Therefore, person re-identification algorithms based on dictionary feature expression were proposed. Liu et al. [13] proposed a Semi-supervised Coupled Dictionary Learning(SSCDL) algorithm, which supervised dictionary learning of samples from gallery and probe respectively, and trained two dictionary learning in same framework to transform original features into mid-level dictionary features. On this basis, Jing et al. [14] conducted a low-rank constraint on feature expression of dictionary learning to improve the robustness and strong discriminative ability of mid-level features.

Compared with the hand-crafted feature, the mid-level features have a stronger feature expression ability. In particular, the filtering features have stronger suitability to illumination and geometric changes. However, due to the complex scenes of person re-identification, the effect of recognition is still unsatisfactory.

2.3 Semantic Feature

The feature of semantic features [15,16,17,18] contain rich semantic information, so it is more robust for changes in illumination, background and spatial distribution of person images. Yang et al. proposed a semantic feature expression method based on color names, combined with the method of saliency learning, used saliency color in person images to express person color features, and converted color information into semantic expression method of color name, making features more robust to changes of dressing color. Layne et al. [16] went a step further and marked 15 attribute features include dressing styles, hairstyles and genders. They discriminated attributes of persons by SVM classifier, and then applied discriminating semantic features to person re-identification. Su et al. [17] obtained low-rank matrix of semantic features by using multi-task learning, and converted semantic features through low-rank matrix to improve discriminative ability of features.

3. Metric Learning Based Method for Person Re-identification

3.1 Formulation of Similarity Metric for Person Re-identification

The task of person re-identification can be converted into the discrimination of the positive and negative pair of person sample images, that is the discrimination that any pair of images belong to the same or different person. The mathematical principle is as follows:

$$d\left(\mathbf{x}_i^A, \mathbf{x}_j^B\right) = \left(\mathbf{x}_i^A - \mathbf{x}_j^B\right)^T \mathbf{M}\left(\mathbf{x}_i^A - \mathbf{x}_j^B\right) \tag{1}$$

Define $\{\mathbf{x}_i^A\}$ and $\{\mathbf{x}_j^B\}$, each representing the feature set of person image obtained from two different monitoring scenes, $\mathbf{x}_i^A$ is called the image feature of the i-th person in scene A and $\mathbf{x}_j^B$ is called the image feature of the j-th person in scene B. $(\mathbf{x}_i^A, \mathbf{x}_j^B)$ forms a sample pair to be identified. lij is label, it is called a positive sample pair if lij=1, while it is called a negative sample pair if lij=0. What we research is measuring the similarity distance between samples.

The measurement of sample similarity is divided into unsupervised measurement models, including the traditional Euclidean distance and the Babbitt distance [21]. However, due to the drastic change of features in person re-identification, unsupervised measurement results are too low in accuracy. Therefore, researchers study supervised measurement model and learn a metric subspace

by using label information, so that the distance of a positive sample pair in the metric space is much smaller than the distance of a negative sample pair. The distance function is as follows:

$$d\left(\mathbf{x}_i^A, \mathbf{x}_j^B\right) = \left(\mathbf{x}_i^A - \mathbf{x}_j^B\right)^T \mathbf{M}\left(\mathbf{x}_i^A - \mathbf{x}_j^B\right) \tag{2}$$

where $\mathbf{M}$ is the metric matrix of Mahalanobis distance.

3.2 Metric Learning based Method for Person Re-identification

Metric learning model is a supervised training method to measure similarity distance between samples by projecting original data into metric subspace. Due to complexity of scenes, traditional methods such as Euclidean distance and Bhattacharyya distance cannot effectively measure similarity of person feature vectors. By constraining the positive and negative sample distances, optimization model of projection matrix of sample feature vector is established. Also, due to strong variance of image features in person re-identification task, diversity of data used for model training is seriously insufficient, resulting in poor model generalization.

Metric learning model is main focus of person re-identification algorithm research after feature model research enters bottleneck, and remarkable progresses have been achieved [3,18,19,20,21,22]. Zheng et al. [18] converted the problem to comparation of relative distance, and built an optimization model by constraining distance of each person in positive samples is far less than negative samples, called Probabilistic relative short comparison (RDC) model. Davis et al. [19] introduced information theory into person re-identification field, proposed the Information-theoretic metric learning (ITML) model. Koestinger et al. [20] proposed probability model of positive and negative samples based on the assumption of zero-mean Gaussian distribution of positive and negative samples. Based on this probability model, a simple and effective statistical inference model KISSME algorithm is proposed. Zhang et al. [21] proposed a Naive Foley-Sammon Transform algorithm, which learned a naive discriminant space based on foley-Sammon Transform. Liao et al. [3] proposed XQDA algorithm based on Fisher discriminant analysis (FDA) algorithm, specific method is shown in section 2.3. Chen et al. [22] designed a Explicit Polynomial Kernel Feature Map model to control impact of strong variation in person image features, and proposed Multi-cue Example-guided Similarity function on Polynomial kernel feature map (MESP) method

3.3 XQDA model for Person Re-identification

Due to the large feature dimensions, the XQDA algorithm aims to learn a subspace and uses the idea of LDA algorithm by supervised learning to minimize the divergence of positive sample pairs and maximum the divergence of negative sample pairs. The distance metric function in the subspace is as follows:

$$d(\mathbf{x}_i^p, \mathbf{x}_j^g) = (\mathbf{x}_j^g - \mathbf{x}_i^p)^T \mathbf{W}\mathbf{H}\mathbf{W}^T(\mathbf{x}_j^g - \mathbf{x}_i^p) \tag{3}$$

where $\mathbf{W} = (w1, w2, \cdots, wr) \in \mathbb{R}^{d \times r}$ represents the subspace projection matrix. The positive and negative samples respectively obey a zero mean Gaussion distribution. The learning of subspace is to solve the optimization problem of Fisher discriminant function to maximize equation 2-4.

$$\max \ J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_E \mathbf{w}}{\mathbf{w}^T \mathbf{S}_I \mathbf{w}} \tag{4}$$

where SE and SI are divergence of extrapersonal and intrapersonal class. Eq. (2-4) can be formulated as Genralized Rayleigh Quotient and converted into solve the eigenvalue of $\lambda \mathbf{w} = \mathbf{S}_I^{-1} \mathbf{S}_E \mathbf{w}$. Sort the eigenvalues from large to small and use top r eigenvectors as the subspace projection matrix W. Finally, learn a metric model in this subspace via KISSME algorithm.

4. Deep Learning Based Method for Person Re-identification

4.1 End-to-end based Deep Framework for Person Re-identification

In deep learning algorithms, convolutional network establishes connection between a wide range of pixels through convolution. Through end-to-end learning mode, feature extraction and metric model are organically combined together, thus elevating the development of computer vision task to a

new height. In early researches on person re-identification algorithms based on deep learning model [23,24,25], most of them took Siamese network as baseline. Li et al. [23] proposed a Deep Filter Pairing Neural Network (FPNN) network model, which uses a weighted sharing two-branched network to extract features from two images from probe and gallery set respectively, and measure the distance between two person images in final full connection layer. However, due to impact of misalignment and change of illumination, person re-identification algorithms base on deep model don't perform well enough.

4.2 Hand-crafted Feature based Deep Framework for Person Re-identification

Since early deep learning-based person re-identification algorithms did not achieve such amazing results in other computer vision fields, people have made various attempts on deep learning-based person re-identification algorithm, among which the combination of manual features and deep network is one of the ideas. Wu et al. [26] proposed a method for integrating ELF manual features with CNN convolution features. In this method, they combined features extracted from convolutional neural network and manual features, used traditional metric model to measure similarity of person samples, and achieved certain effect. Li et al. [27] also proposed a feature expression model Multi-Statistics Cascade on Pyramid (MSCP) combining manual feature and deep feature. This model combines features of deep PCA network feature and LOMO[3] to improve the accuracy to some extent.

4.3 Hybrid Deep Learning Method for Person Re-identification

Hybrid Deep Learning Based Method combines advantages of deep learning and metric learning. This kind of models take advantage of strong feature expression ability and robustness of the deep network model, as well as strong discriminative ability of metric model. Firstly, deep model is used to extract feature of deep expression, and then deep feature is substituted into metric model for person classification and recognition. Xiao et al. [28] proposed a domain-guided person image feature expression model, which put multiple datasets together to train feature expression model, and then designed a Dropout method for each dataset. Each dataset was separately substituted into the model to guide the Dropout training method and extract a more effective feature. Finally, extracted deep features were substituted into KISSME metric model for metric learning. Zheng et al. [29] proposed a PoseBox Fusion (PBF) CNN, which introduces pose-invariant embeddinghas and effectively solved impact of misalignment in person re-identification. Through a PoseBox structure, the model aims persons with different poses to a standard pose, which effectively improves recognition accuracy.

In addition, He et al. [30] proposed a person re-identification model based on residual network, this model explicitly defines deep network layers to study residual function. Using input layer as reference, rather than learning function without reference, the method uses residual network converges faster, thus improves performance of network model. Zheng et al. [31], taking human body structure as a prior knowledge, divided images into patches, and added multi-label on each patch according to body structure. Through training of CNN network model with multi-label, the model performs better in feature extraction and recognition.

5. Datasets for Person Re-identification and Experiments

5.1 Publicly Datasets for Person Re-identification

As person re-identification attracts more attention, researchers provide more person image datasets for person re-identification research. The commonly used data sets and relevant parameters are shown in Table 1 below:

Table 1 Introduction of common person re-identification datasets

Datasets	Year	Image/Video	#ID	#Image	#Cam
VIPeR	2007	Image	632	1264	2
GRID	2009	Image	250	1,275	8

CUHK01	2012	Image	971	3,884	2
CUHK03	2014	Image	1,467	13,164	2
iLIDS-VID	2014	Video	300	600(44k)	2
MARS	2016	Video	1261	20,715(1M)	6

VIPeR, CUHK01, CUHK03 and GRID are all image-based person re-identification datasets, and VIPeR is one of the earliest the most widely used person re-identification datasets. This dataset is aimed at person re-identification in a single-shot scenario. It contains 1264 image samples of 632 persons, that is, Each person has one and only one image under each camera. The image samples in this dataset include lighting, shooting angle, pose and other changes, which are very challenging. CUHK01 and CUHK03 are open datasets provided by The Chinese University of Hong Kong in 2012 and 2014 respectively. CUHK01 dataset is collected by manual annotation. This dataset contains 971 person image samples captured under two cameras, and is used for person re-identification in multi-shot scenes. Each person is represented by two images under each camera. The images in this dataset have high resolution and clarity, and all person images are normalized to a 160x60 image resolution. CUHK03 is also a person re-identification image dataset for multi-shot. This dataset is one of the largest person re-identification datasets, containing 13,164 pictures of 1467 persons. It is worth noting that the number of images per person in the dataset is not fixed. Moreover, the dataset provides two kinds of data: manual annotation and algorithm annotation. GRID dataset is one of the most difficult person re-identification datasets. It not only has poor image quality, but also has complex scene and complex identification environment. The dataset provides 250 person image samples, each person contains two images, in addition, the dataset also provides 775 person images to confuse identification, increasing the difficulty of identification.

iLIDIS-VID and MARS datasets are video-based person re-identification datasets. iLIDIS-VID is a video person re-identification dataset provided in 2014, which is more neat and challenging than the earlier video-based person re-identification datasets. The dataset is video data obtained from the CCTV surveillance network in the airport lobby. This dataset contains 600 video images of 300 persons, in which there are serious similarities in clothing, changes in lighting and view points, complex background and occlusion problems. The MARS video person re-identification dataset is an expanded version of the Market-1501 dataset, with the number of images expanded from 32668 to 1191,003.

5.2 Comparation and Analysis of Models

5.2.1 Comparation and Analysis of Existing Methods

The performance evaluation of person re-identification algorithm is one of the key tasks of person re-identification. Different from the traditional recognition task in the field of computer vision, in the application scenario of person re-identification task, we not only focus on the most similar samples, but also focus on multiple recognition results with high similarity ranking. Currently, CMC(Cumulative Match Characterisric) curve is applied as the performance evaluation index of person re-identification algorithm. The corresponding calculation formula is as follows:

$$CMC(l) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(rank(P_i) < l) \tag{5}$$

where l is the rank of CMC cumulative accuracy rank=l, that is, the test distance is l ranked from small to large, and N is the number of gallery samples in the test samples. $\mathbb{I}(\cdot)$ is a symbolic function, that is, when the internal variable of the function is true, the corresponding value of the function is 1; otherwise, it is 0. rank($\cdot$) represents the ranking calculation of sample distance, Pi is the positive sample distance of the i-th gallery sample. rank(Pi) represents the ordering of its positive samples.

As can be seen from the definition of this index, CMC curve reflects the probability of finding the correct result among the first k matching results. Fig. 4-1 shows the recognition results of different algorithms on VIPeR dataset. When rank=1, it represents the first-round recognition rate, that is, the classification accuracy in the traditional recognition task. When rank=n, the corresponding value represents the ratio of correct results in the first N images.

In addition to the CMC curve, people often use another more direct measurement form, Rank-N table. The accuracy rate of key matching points in CMC curve is given in the table. The commonly used ones are rank-1, rank-5, rank-10 and rank-20. Where, rank-5 represents the probability of the first five images that can match the correct person. The higher value represents the better the effect of algorithm.

5.2.2 Semi-supervised Metric Learning Based on Deep Features

In view of the poor generalizability of traditional measurement methods and the inrobustness of complex change features, this paper introduces the idea of semi-supervision on the basis of XQDA algorithm and uses JSTL to extract depth features for metric learning to improve the recognition accuracy. Firstly, the positive and negative sample pairs $(\mathbf{x}_i^p, \mathbf{x}_j^g)$ are randomly divided into training samples $X = \left\{\mathbf{x}_i^p, \mathbf{x}_j^g, l_{ij}\right\}$ and test samples $Y = \left\{\mathbf{y}_k^p, \mathbf{y}_s^g, l_{ks}\right\}$, $l_{ij}, l_{ks} = 0,1$ respectively represent the training set label and test set label. XQDA metric learning algorithm is used to train the training samples to get the initial metric model (W,M), and then the similarity metric is performed on the test samples with this model to get the metric matrix $\mathbf{D}_1 = \left\{\mathbf{d}_{1,1}, \cdots, \mathbf{d}_{1,k}, \cdots, \mathbf{d}_{1,n}\right\}$, where the $\mathbf{d}_{1,k} = \left\{d_{k,1}, d_{k,2}, \cdots, d_{k,n}\right\}$ represents the distance between the k-th target and each sample in the gallery collection. Sort the distance $\mathbf{d}_{1,k}$ from small to large and get $\mathbf{d}'_{1,k} = \left\{d_{k,m_1}, d_{k,m_2}, \cdots, d_{k,m_n}\right\}$, $d_{k,m_1} \le d_{k,m_2} \le \cdots \le d_{k,m_n}$. $\left(\mathbf{y}_k^p, \mathbf{y}_{m_1}^g\right)$ compose a positive sample pair, that is, $l'_{km_1} = 1$. Then, choose any sample $\left\{\mathbf{y}_{m_s}^g \mid m_s \neq m_1\right\}$ and compose a negative sample pair, that is, $l'_{km_s} = 0$. The lebel set is $\left\{l'_{ks}\right\}$. Using the label information, take test samples into the XQDA model to learn a subspace. The method is as follows:

$$\max \quad J(\mathbf{w}) = \frac{\mathbf{w}^T S'_B \mathbf{w}}{\mathbf{w}^T S'_I \mathbf{w}} = \frac{\mathbf{w}^T \frac{1}{n_B} \sum_{k=1}^{n_B} \left(\mathbf{y}_k^p - \mathbf{y}_{m_1}^g\right)\left(\mathbf{y}_k^p - \mathbf{y}_{m_1}^g\right)^T \mathbf{w}}{\mathbf{w}^T \frac{1}{n_I} \sum_{k=1}^{n_B} \left(\mathbf{y}_k^p - \mathbf{y}_{m_s}^g\right)\left(\mathbf{y}_k^p - \mathbf{y}_{m_s}^g\right)^T \mathbf{w}} \tag{6}$$

where $n_B = n_I$ respectively represents number of positive and negative sample pairs marked by pseudo label information in semi-supervised learning. Use generalized Lagrange multiplier method to solve the above problem, and obtain new projection matrix W1.

Based on the above method, a new subspace (W1,M1) is trained. Then, a new metric matrix $\mathbf{D}_2$ is obtained by measuring the test samples in subspace (W,M) and (W1,M1). Using the distance information in $\mathbf{D}_2$, a new set of labels $\left\{l''_{ks}\right\}$ is calculated, and a new subspace (W2,M2) is obtained according to the training of formula (4-2). By repeating the above process, we can get multiple set of subspaces $\left\{\left(\mathbf{W}_k, \mathbf{M}_k\right) \mid k = 1, 2, \cdots, n\right\}$.

Based on a set of subspaces and original subspace trained by the above semi-supervised method, the joint metric model of samples is built, as shown in formula 4-3 :

$$d\left(\mathbf{y}_k^p, \mathbf{y}_s^g\right) = d_{(\mathbf{W},\mathbf{M})}\left(\mathbf{y}_k^p, \mathbf{y}_s^g\right) + \sum_{i=1}^{c} \alpha_i d_{(\mathbf{W}_i,\mathbf{M}_i)}\left(\mathbf{y}_k^p, \mathbf{y}_s^g\right) \tag{7}$$

where α_i is the equilibrium coefficient.

Based on the above semi-supervised metric learning model, the test set data were identified and the final CMC curve was calculated. The results are shown in Fig. 1.

Table 2 The cumulative matching scores (%) with the VIPeR database(P=316) at rank 1,5,10,20 are listed.

Method	rank=1	rank=5	rank=10	rank=20
LMNN[32]	25.63	56.96	71.20	85.44
MVSLD2L[34]	20.79	45.08	61.24	81.36
Improved Deep[24]	34.81	63.50	75.00	80.00
KCCA[33]	30.16	62.69	80.51	91.80
XQDA+LOMO[3]	40.00	68.26	80.51	91.80
NFST-LOMO[37]	42.28	71.46	82.94	92.06
MLAPG+LOMO[36]	38.92	70.32	82.28	93.04
KISSME+LOMO[20]	34.14	64.81	77.63	89.40
TCP[35]	47.8	74.7	84.8	91.1
Ours-JSTL	50.35	71.14	82.31	92.06

From Table 4 and Fig. 1 it can be observed that out method improves the performance at rank 1-100 from the original XQDA algorithm. Especially, a 10.35% performance gain can be obtained for the rank-1 accuracy. It can be seen that semi-supervised learning method can effectively improve the problem of weak generalization in person re-identification.
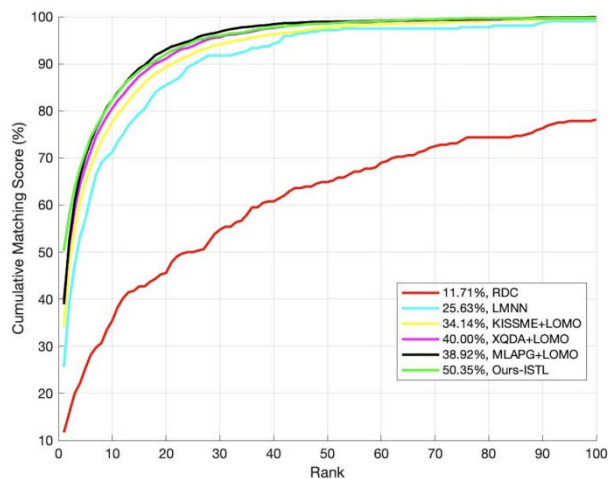

Figure 1 CMC curves on different methos with the VIPeR database(P=316).

6. Conclusion

Person re-identification task is the core difficulty in the field of computer vision, and it is the extension of the three recognition tasks in the field of computer vision in complex scenes. At present, the research of person re-identification method is still in the laboratory stage, which is far from practical application, and its accuracy needs to be further improved. The difficulty of person re-identification task lies in the complexity of feature change, especially the change of spatial distribution of person in the image caused by inaccuracy. In recent years, with the introduction of deep learning, person re-identification accuracy has been greatly improved. Based on the convolutional neural network method in deep learning, more pixels in the space are connected together, making feature extraction more robust. At present, the research on person re-identification is still based on the dataset in the ideal environment without considering the occlusion problem. The recognition task in dense environment needs further research before it can be put into practical application.

References

[1] Zheng Z, Zheng L, Yang Y. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro [J]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017:3774-3782.

[2] Bazzani L, Cristani M, Murino V. Symmetry-driven accumulation of local features for human characterization and re-identification. Computer Vision and Image Understanding, 2013, 117(2): 130-144.

[3] Liao S C, Hu Y, Zhu X Y, Li S Z. Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015. 2197-2206.

[4] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Proceedings of the 10th European Conference on Computer Vision. 2008. 262-275.

[5] Farenzena M, Bazzani L, Perina A, Murino V, Cristani M. Person re-identification by symmetry-driven accumulation of local features. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. 2010. 2360-2367.

[6] Gheissari N, Sebastian T B, Hartley R. Person reidentification using spatiotemporal appearance. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006. 1528-1535.

[7] Zeng M Y, Wu Z M, Tian C, Zhang L, Hu L. Efficient person re-identification by hybrid spatiogram and covariance descriptor. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015. 48-56.

[8] Matsukawa T, Okabe T, Suzuki E, Sato Y. Hierarchical Gaussian descriptor for person re-identification. IEEE Conference on Computer Vision and Pattern Recognition. 2016. 1363-1372.

[9] Zhao R, Ouyang W L, Wang X G. Person re-identification by saliency learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 356-370.

[10] Kviatkovsky I, Adam A, Rivlin E. Color invariants for person reidentification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(7): 1622-1634.

[11] Zhao R, Ouyang W L, Wang X G. Learning mid-level filters for person re-identification. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014. 144-151.

[12] Wang S , Zhang D , Liang Y , et al. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[13] Liu X, Song M, Tao D, et al. Semi-supervised Coupled Dictionary Learning for Person Re-identification. IEEE Conference on Computer Vision & Pattern Recognition. 2014.

[14] Jing X Y, Zhu X, Wu F, et al. Super-Resolution Person Re-Identification With Semi-Coupled Low-Rank Discriminant Dictionary Learning. IEEE Transactions on Image Processing, 2017, 26(3):1363-1378.

[15] Yang Y, Yang J, Yan J, Liao S, Yi D, and Li S Z, "Salient color names for person re-identification," in European Conference on Computer Vision, 2014, pp. 536–551.

[16] Layne R, Hospedales T M, Gong S G. Person re-identification by attributes. In: Proceedings of the 2012 British Machine Vision Conference. 2012.

[17] Su C, Yang F, Zhang S L, Tian Q, Davis L S, Gao W. Multitask learning with low rank attribute embedding for person re-identification. IEEE International Conference on Computer Vision. 2015. 3739-3747.

[18] Zheng W S, Gong S G, Xiang T. Person re-identification by probabilistic relative distance comparison. IEEE Conference on Computer Vision and Pattern Recognition. 2011. 649-656.

[19] Davis J V, Kulis B, Jain P, Sra S, Dhillon I S. Information-theoretic metric learning. In: Proceedings of the 24th International Conference on Machine Learning. 2007. 209-216

[20] Koestinger M, Hirzer M, Wohlhart P, Roth P M, Bischof H. Large scale metric learning from equivalence constraints. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012. 2288-2295.

[21] Zhang L, Xiang T, and Gong S, "Learning a discriminative null space for person re-identification," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1239–1248.

[22] D. Chen, Z. Yuan, J. Wang, B. Chen, G. Hua, and N. Zheng, "Exemplar guided similarity learning on polynomial kernel feature map for person re-identification," International Journal of Computer Vision, vol. 123, no. 3, pp. 392–414, 2017.

[23] Li W, Zhao R, Xiao T, Wang X G. DeepReID: deep filter pairing neural network for person re-identification. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014. 152-159.

[24] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015. 3908-3916.

[25] McLaughlin N, Martinez Del Rincon J, Miller P. Recurrent convolutional network for video-based person reidentification. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016. 1325-1334.

[26] Wu S X, Chen Y C, Li X, Wu A C, You J J, Zheng W S. An enhanced deep feature representation for person re-identification. In: Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision. 2016. 1-8

[27] Li Y J, Zhuo L, Hu X C, Zhang J. A combined feature representation of deep feature and hand-crafted features for person re-identification. In: Proceedings of the 2016 International Conference on Progress in Informatics and Computing. 2016. 224-227

[28] Xiao T , Li H , Ouyang W , et al. Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1249-1258.

[29] Zheng L, Huang Y J, Lu H C, Yang Y. Pose invariant embedding for deep person re-identification. arXiv preprint, arXiv: 1701.07732, 2017.

[30] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016. 770-778

[31] Zheng L, Zhang H H, Sun S Y, Chandraker M, Yang Y, Tian Q. Person re-identification in the wild. arXiv preprint, arXiv: 1604.02531, 2016.

[32] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," Journal of Machine Learning Research, vol. 10, no. 1, pp. 207–244, 2009.

[33] G. Lisanti, I. Masi, and A. Del Bimbo, "Matching people across camera views using kernel canonical correlation analysis," in International Conference on Distributed Smart Cameras, 2014, pp. 1–6.

[34] X. Y. Jing, X. Zhu, F. Wu, R. Hu, X. You, Y. Wang, H. Feng, and J. Y. Yang, "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," IEEE Transactions on Image Processing, vol. 26, no. 3, pp. 1363–1378, 2017.

[35] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person reidentification by multi-channel parts-based cnn with improved triplet loss function," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1335–1344.

[36] S. Liao and S. Z. Li, "Efficient psd constrained asymmetric metric learning for person re-identification," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3685–3693.

[37] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1239–1248.