Improved Statistical Interference Model for Person Re-identification

Linxuan Li
Shenzhen University, China

Abstract. Person Re-identification problem is an important and challenging task in computer vision task. Due to the drastic appearance variation caused by misalignment and illumination changing, traditional metric models are failed in similarity measure of pedestrian images. In this paper, a novel metric learning based method is proposed. It establishes a probability inference model based on the probability models of positive pairs and negative pairs. And a balance parameter is proposed in the metric model to deal with the imbalance problem of samples. Finally, experiments are conducted on the VIPeR dataset compared with some metric learning based model. And the test results verified the effectiveness of the proposed model.

Keywords: Person re-identification, statistical interference, metric learning.

1. Introduction

Person re-identification[1-5] is a branch of computer vision, which uses computer vision technology to determine whether there are specific person in images or video sequences. It mainly describes a person who is captured in one view when passing through a non - overlay public area, which can be identified in another view, though the spatial position has changed. Person re-identification is of great value in public security and image retrieval. It can also help mobile phone users to classify their album. Therefore, by combining with large data technology, it can help retailers or business operators to obtain effective customer information. However, the appearance of pedestrians is easily affected by clothing, occlusion, posture and viewpoint. And also because of the lower camera pixels and other hardware constraints, the existing person re-identification technology cannot achieve the desired matching rate.

To track the same person through different camera views located at different physical sites, person re-identification contains two major works: (1)feature representation, is a kind of method that convert the original feature into a set of features that have obvious physical meaning (Gabor, geometric features [corners, invariants], texture [LBP HOG]) or statistical meaning or kernel to highlight representative characteristics under specific space; (2) metric learning, is to learn a distance metric for the input space of data from a given collection of pair of similar/dissimilar points that preserves the distance relation among the training data.

After the great achievement that computer vision researchers have made, that is still a challenge because it requests matching the same person when illumination, pose or viewpoint has changed completely. Even people in the similar clothes can also produce great effect on result. Existing approaches focus on developing discriminative feature representations that are robust against the view/pose/illumination/background changes, or learning a distance metric which can maximize the inter-class variations and minimize the intra-class variations, or both jointly.

Usually, the description based on feature is used to compare similarity with standard distance. However, when the same person cross multiple cameras without overlapping area, different appearance characteristics are affected by the viewpoint, illumination and other factors. Standard distance measures treat each feature equally, without ignoring features that are ineffective when used independently. Therefore, we can project people features into a new distance metric space by distance learning, so that the distance between different images of the same pedestrian is smaller than the distance between different people. Distance learning method is ordinarily based on Mahalanobis distance. By learning a projection matrix, we can maximize the inter-class variations and minimize the intra-class variations. Unlike simple similarity, it includes the training process for the identified

samples, and maps the eigenvectors to a more easily distinguished space by learning the measure matrix.

In fact, both of them are two independent steps in a complete process of person re-recognition. By studying these two directions separately, we can find their own solutions, and then combine any two of them together, select a satisfactory combination and improve it, so as to suit our purposes.

2. Related Work

The most advanced feature sets are chiefly based on region or block feature models, while other effective features are expressed by combining multiple types of features. For example: Symmetric Driven Local Cumulative Feature (SDALF)[6], Custom Patterns Structural Feature (CPS)[7] and Local Fisher Vector Coded Hybrid Feature (ELDFV)[8]. Although these feature models have made remarkable improvements in image representation and distinguishing, the changes of person image features are too complex to improve the final recognition accuracy. Yang[9] and others proposed an image color feature representation model based on color name, which makes the feature representation more robust to illumination change and local geometric transformation. The SCNCD feature also has a good recognition accuracy on this problem. Liao Shengcai[10] and others fully extracted the appearance information of the image through the combination of color histogram and texture histogram, forming a more complex feature description model LOMO model. Based on the feature extraction method of the model, an ultra-high dimension feature vector is constructed to calculate the local color and texture distribution, and has obtained the outstanding result. It is widely used in person re-identification tasks. However, the improvement of person re-identification method based only on feature model is limited, and the requirement of high-dimensional features on the ability of measurement algorithm is higher.

To against illumination and viewpoint changes, a number of methods have been proposed. Kilian Q. Weinberger[11] and all have proposed Distance Metric Learning for Large Margin Nearest Neighbor Classification (LMNN), prompting the k-nearest neighbors always belong to the same class while examples from different classes are separated by a large margin as result of learning a Mahanalobis distance metric for k-nearest neighbor (kNN) classification by semidefinite programming. Martin[12] have proposed Large Scale Metric Learning from Equivalence Constraints (KISSME), on scalability and the required degree of supervision of existing Mahalanobis metric learning methods. Wei-Shi Zheng[13] has proposed relative distance comparison (RDC), introducing a novel relative distance comparison model which is formulated to maximize the likelihood of a pair of true matches having a relatively smaller distance than that of a wrong match pair in a soft discriminant manner. In addition, RDC algorithm is improved by probability model, and PRDC[14] algorithm is proposed. Shengcai Liao[10] have proposed an effective feature representation called Local Maximal Occurrence (LOMO) to analyze the horizontal occurrence of local features, and a subspace and metric learning method called Cross-view Quadratic Discriminant Analysis (XQDA) to learn a discriminant metric. Davis[15] and others introduced the method of information theory and proposed an ITML algorithm for learning similarity distance metric matrix.

3. Method

The purpose of person re-identification task is to measure the similarity of pedestrian captured by different cameras in the surveillance network, so as to locate and track specific pedestrian. Therefore, the key to re-identification is to establish effective distance or similarity measure function.

Given dataset $\mathbf{X}=\left\{\mathbf{x}_i^p,\mathbf{x}_j^g\right\};\mathbf{x}_i^p,\mathbf{x}_j^g\in\mathbb{R}^{2d\times N}$, $\mathbf{x}_i^p,\mathbf{x}_j^g$ represents pedestrian images captured under two different cameras. Let l_{ij} denote the label information of the sample to the relationship between the two samples $\left(\mathbf{x}_i^p,\mathbf{x}_j^g\right)$. If $l_{ij}=1$, $\left(\mathbf{x}_i^p,\mathbf{x}_j^g\right)$ is a positive sample pair, the two image corresponds to the same pedestrian; if $l_{ij}=0$, $\left(\mathbf{x}_i^p,\mathbf{x}_j^g\right)$ is a negative sample pair, the two image corresponds to the different pedestrian. The essence of pedestrian recognition task is to match every $\mathbf{x}_i^p$ in the scene of

camera A to the samples in sets $\{\mathbf{x}_j^g\}$ captured by camera B. Calculate the distance between each pair of images, find the nearest sample to $\mathbf{x}_i^p$.

However, the method of sample similarity measurement based on Euclidean distance is too simple. The appearance features of pedestrian images change dramatically in the process of crossing the monitoring scene, resulting in the recognition accuracy unsatisfactory. Metric learning based method is to learn a Mahalanobis distance. Therefore, the distance measure function of sample pairs is expressed as follows:

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

The metric matrix $\mathbf{M}$ is a semi-positive definite matrix. Let $\mathbf{o}_{ij} = \mathbf{x}_i^p - \mathbf{x}_j^g$, $\mathbf{O}_+ = \{\mathbf{o}_{ij}\}$ represents the positive sample pair set, and $\mathbf{O}_- = \{\mathbf{o}_{ij}\}$ represents the negative sample pairs set. Based on the population distribution hypothesis of positive and negative sample pairs, a Mahalanobis distance is learned in feature space to improve the effect of sample similarity measurement. From the point of view of statistical inference, the optimal decision-making scheme for the similarity of samples to $\left(\mathbf{x}_i^p, \mathbf{x}_j^g\right)$ can be solved by the probability model of the distribution of samples. The positive pair's probability function is $p_0(\mathbf{x}_i^p, \mathbf{x}_j^g) = \frac{1}{\sqrt{2\pi\left|\sum_{l_{ij}=0}\right|}} \exp(-1/2\, \mathbf{o}_{ij}^T \sum_{l_{ij}=0}^{-1} \mathbf{o}_{ij})$ and the negative pair's probability function is $p_1(\mathbf{x}_i^p, \mathbf{x}_j^g) = \frac{1}{\sqrt{2\pi\left|\sum_{l_{ij}=1}\right|}} \exp(-1/2\, \mathbf{o}_{ij}^T \sum_{l_{ij}=1}^{-1} \mathbf{o}_{ij})$. $\sum_{l_{ij}=1}^{-1}$ and $\sum_{l_{ij}=0}^{-1}$ represent the covariance matrix of the positive and negative sample pairs in the projection space, respectively. $p_0(\mathbf{x}_i^p, \mathbf{x}_j^g)$ denotes the probability of negative sample pairs; $p_1(\mathbf{x}_i^p, \mathbf{x}_j^g)$ denotes the probability of positive sample pairs. For the distance between the sample pairs, the greater the difference, the lower the probability of belonging to the positive sample. Therefore, we define the sample distance function as follows:

$$d(\mathbf{x}_i^p, \mathbf{x}_j^g) = \gamma\, p_0(\mathbf{x}_i^p, \mathbf{x}_j^g) - p_1(\mathbf{x}_i^p, \mathbf{x}_j^g) \tag{2}$$

γ is the balance coefficient, which can effectively solve the problem of poor model resolution caused by the imbalance of the logarithmic number of positive and negative samples. The distance model (1) is simplified. First, the formula (2) is taken logarithm on the right, and got the formula (3):

$$d(\mathbf{x}_i^p, \mathbf{x}_j^g) = \frac{1}{2}\mathbf{o}_{ij}^T \sum_{l_{ij}=1}^{-1} \mathbf{o}_{ij} + \ln\left(\sqrt{2\pi\left|\sum_{l_{ij}=1}\right|}\right) - \frac{\gamma}{2}\mathbf{o}_{ij}^T \sum_{l_{ij}=0}^{-1} \mathbf{o}_{ij} - \ln\left(\sqrt{2\pi\left|\sum_{l_{ij}=0}\right|}\right) \tag{3}$$

The constant term in the above formula makes the similarity between samples shift to a certain extent on the whole, and does not affect the comparison results of the samples. Therefore, the constant term in the above formula is removed and the following formula (4) is obtained:

$$d(\mathbf{x}_i^p, \mathbf{x}_j^g) = \mathbf{o}_{ij}^T \left(\sum_{l_{ij}=1}^{-1} - \gamma \sum_{l_{ij}=0}^{-1}\right) \mathbf{o}_{ij} \tag{4}$$

Finally, the Mahalanobis distance measurement model is as follows:

$$d(\mathbf{x}_i^p, \mathbf{x}_j^g) = \left(\mathbf{x}_i^p - \mathbf{x}_j^g\right)^T \mathbf{M} \left(\mathbf{x}_i^p - \mathbf{x}_j^g\right) \tag{5}$$

Where $\mathbf{M} = \left(\sum_{l_{ij}=1}^{-1} - \gamma \sum_{l_{ij}=0}^{-1}\right)$. The mathematical expressions of $\sum_{l_{ij}=1}^{-1}$ and $\sum_{l_{ij}=0}^{-1}$ are as follows:

$$\Sigma_{l_{ij}=1}^{-1} = \frac{1}{n_1}\sum_{l_{ij}=1}(\mathbf{x}_i^p - \mathbf{x}_j^g)(\mathbf{x}_i^p - \mathbf{x}_j^g)^T; \Sigma_{l_{ij}=0}^{-1} = \frac{1}{n_2}\sum_{l_{ij}=0}(\mathbf{x}_i^p - \mathbf{x}_j^g)(\mathbf{x}_i^p - \mathbf{x}_j^g)^T \quad (6)$$

4. Experiment

In order to verify the effectiveness of this method, the most widely used VIPeR database is selected to test the algorithm. Firstly, in 1, the database and simulation experiment settings are introduced; secondly, the evaluation index is introduced in 2; finally, the experimental results and analysis are given in 3. The VIPeR [3] dataset is the earliest publicly available image database for person re-identification. The database contains 632 pedestrian images captured in two different outdoor monitoring scenes. 632 pedestrians correspond to one image in two scenes, so the database contains 1264 pedestrian images. The resolution of the image is low and its size is 128 * 48 pixel size.

In this paper, the cumulative accuracy curve (CMC curve) [5,13] is selected as the evaluation index of the recognition accuracy of the algorithm. The calculation method is as follows:

$$CMC(l) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(rank(P_i) < l) \quad (7)$$

Among them, l represents rank=l of cumulative accuracy of CMC, which means the test distance is ranked l according to the rank from small to large, and N is the number of gallery samples in the test sample. For the sign function, that is, the inner variable of the function is true, the corresponding function value is 1, otherwise it is 0. Represents the sample distance ranking calculation, which is the positive sample distance of the first gallery sample and the ranking of its positive samples. As shown in Figure 1 below, the test results based on VIPeR database are presented. The algorithm and the CMC curves of the recognition accuracy of the four comparison algorithms are given. In order to ensure the validity of the experimental results, the experimental results presented in this paper are the average recognition accuracy of 10 repetitive independent experiments. The abscissa represents the rank number, and the recognition accuracy of the ordinate represents the correct sample ratio in the recognition results corresponding to the ranking of the first n. From the results shown in the figure, it can be seen that the recognition accuracy of the algorithm in this paper achieves the best recognition accuracy among all the comparison algorithms. In order to understand the performance of the algorithm more clearly, the relevant data are counted, as shown in Table 1.

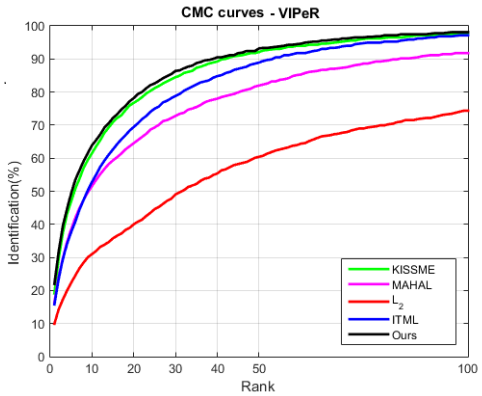

Fig. 1 Experiments on VIPeR compared with methods

Table 1 is the statistical table of the identification accuracy of the compared experiments. The rank-1, rank-5, rank-10 and rank-20 identification rates are given in the table, which show the proportion of correct recognition results in the top 1, 5, 10 and 20 recognition results, respectively. From the results in the table, we can clearly see that the proposed method in this paper achieves the best recognition accuracy in all the indicators. Especially in the rank-1 recognition accuracy, it improves the best comparison algorithm KISSME by 4.07%.

Table 1. Identification accuracy statistical results

methods	p=316			
	r=1	r=5	r=10	r=20
MAHAL	16.61	38.30	51.20	64.72
IDENTITY	10.15	22.68	31.43	40.77
ITML	15.46	38.32	53.06	69.68
KISSME	19.35	47.89	62.77	**77.21**
Ours	**242**	**50.00**	**63.92**	76.96

Furthermore, the sensitivity of parameter γ is also tested on the VIPeR over different values (see in Fig.2). As shown in Fig. 2, the rank-1, rank-5, rank-10 and rank-20 identification rate curves are given under different values. The identification rates of proposed method are first increase and then decrease along with the increase of γ 's value and reach the best performance when γ =1.2. Therefore, we set γ =1.2 in this paper.

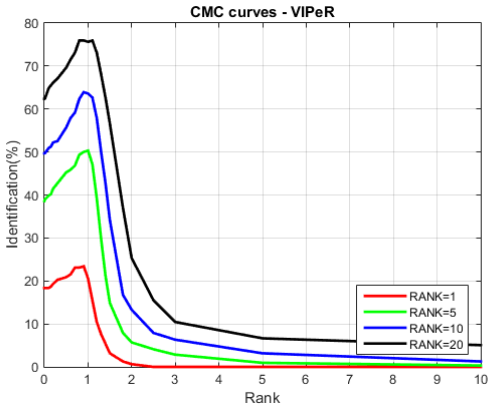

Fig. 2 Sensitivity analysis of parameter γ

5. Conclusion

Person re-identification task is matching the pedestrian image captured in one camera view with pedestrian images captured in other camera views. Metric learning based method is trying to learn a distance metric model to measure the similarity of image pairs. While person re-identification is suffering drastic appearance variation problem and imbalance of samples problem. In this paper, we propose a novel probability inference model to learn a metric model by weighting the probabilities of positive pairs and negative pairs with different values.

References

[1]. Liu X., Tao D., Song M., and Zhang L., Bu J. and Chen C. Learning to Track Multiple Targets[J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, 26(5):1060.

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

[3]. Gray D., Brennan S., and Tao H. Evaluating appearance models for recognition, reacquisition, and tracking[C]// in IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2007:1-7.

[4]. Zhang Z, Chen Y, Saligrama V. A Novel Visual Word Co-occurrence Model for Person Re-identification[C]// Workshop at the European Conference on Computer Vision. Springer International Publishing, 2014:122-133.

[5]. Hirzer M., Roth P. M., Kostinger M., and Bischof, H. Relaxed pairwise learned metric for person re-identification[C]// European Conference on Computer Vision, Florence, 2012: 780-793.

[6]. Bazzani L., Cristani M., Murino V. (2014) SDALF: Modeling Human Appearance with Symmetry-Driven Accumulation of Local Features. In: Gong S., Cristani M., Yan S., Loy C. (eds) Person Re-Identification. Advances in Computer Vision and Pattern Recognition. Springer, London.

[7]. Dong S C, Cristani M, Stoppa M, et al. Custom Pictorial Structures for Re-identification [C]. British Machine Vision Conference, 2011: 6.

[8]. Ma B., Su Y., Jurie F. (2012) Local Descriptors Encoded by Fisher Vectors for Person Re-identification. In: Fusiello A., Murino V., Cucchiara R. (eds) Computer Vision – ECCV 2012. Workshops and Demonstrations. ECCV 2012. Lecture Notes in Computer Science, vol 7583. Springer, Berlin, Heidelberg.

[9]. Yang Y., Yang J., Yan J., Liao S., Yi D., Li S.Z. (2014) Salient Color Names for Person Re-identification. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham.

[10]. Shengcai Liao, Yang Hu, Xiangyu Zhu, Stan Z. Li, "Person Re-Identification by Local Maximal Occurrence Representation and Metric Learning", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2197-2206.

[11]. K. Q. Weinberger, L. K. Saul, Distance metric learning for large margin nearest neighbor classification, Journal of Machine Learning Research 10 (1) (2009) 207–244.

[12]. Roth P.M., Hirzer M., Köstinger M., Beleznai C., Bischof H. (2014) Mahalanobis Distance Learning for Person Re-identification. In: Gong S., Cristani M., Yan S., Loy C. (eds) Person Re-Identification. Advances in Computer Vision and Pattern Recognition. Springer, London.

[13]. Wei-Shi Zheng, Shaogang Gong and Tao Xiang, "Reidentification by Relative Distance Comparison", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, pp653 – 668, 26 June 2012.

[14]. Wei-Shi Zheng, Shaogang Gong, Tao Xiang, "Person re-identification by probabilistic relative distance comparison", Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2011.

[15]. Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, Inderjit S. Dhillon, "Information-theoretic metric learning", ICML '07 Proceedings of the 24th international conference on Machine learning, pp. 209-216, 2007.