International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 1
January 2018

Effectiveness of the Proposed Metric Learning Method for Person Re-Identification

Nidhi Agarwal, Aditya Gupta, Adarsh, Abhinav Tyagi, Alok Vajpai
(IPEC, GHAZIABAD)

Abstract: Person of interest is a term used by law enforcement when identifying someone involved in a criminal investigation who has not been arrested or formally accused of a crime. Person re-identification is the process of finding a person of interest in a concerned suspected area of crime bounded by cameras. In this paper, a new method for person re- identification for the similarity learning method is proposed. Similarity learning is closely related to distance metric learning. Metric learning is the task of learning a distance function over objects. Already existing metric learning methods are usually dependent on linear transformation by using pairwise or triplet sparse constraints. But the disadvantage of using this method is that since many negative pairs and triplets with matching conditions are discarded, the discriminative information is unable to be retrieved properly. Also, we introduce a nonlinear transformation function for metric learning to extract the actual feature space values and judge for internal product similarity using the structured learning schema. We used the VIPeR dataset and our proposed method has achieved quite satisfactory results on these datasets.

Index Terms—Empirical evaluation, metric learning, person of interest, person re-identification.

I. INTRODUCTION

PERSON re-identification refers to the process of matching a selected set of image dimensions with a set of gallery images. It has gained a lot of concern in recent years ([1]– [8]) as being part of its applications in video surveillance, e.g. cross-camera tracking, multi-camera analysis and pedestrian movement search. Already existing methods used for person re-identification are broadly categorized into two types: image descriptor based and similarity/distance metric learning based. Various methods in the first type are broadly engaged in designing or learning image descriptors which are important for illumination and various viewpoint changes. Various effective person image descriptor methods are proposed. However, it is mostly difficult to construct descriptors concerning mainly the criteria of robustness against various changes caused by lightings effects, body poses, viewpoints, but also discriminant power against different identities. The re-identification process is further improved by using metric learning methods in addition to constructor descriptors.

Considering the above solutions in mind, another category is discussed in this paper where metric for distance calculation is selected to effectively measure the similarities and differences of person images. All the metric learning methods used till now ([12]– [15]) are based on a Mahalanobis distance by employing pairwise or triplet constraints. With the aim to balance the number of positive and negative matching pairs, while also trying to control the cost of computation, only a small number of training pairs or triplets are used for learning ([13], [14], [16].

Also, the conventional metric learning methods follow a linear transformation function for the actual feature space, which most of the times alone may not be sufficient enough to capture the various nonlinear values where the person images usually fall [19]. To fit the data distribution in a better way, some more methods like local metric learning ([20], [21]) and multi-task metric learning [7] are also proposed. In addition to this, some nonlinear embedding techniques are also proposed for person re-identification ([22], [23]), RGB-D sensor-based classification [24], saliency detection [25]. In this paper, the features are explicitly embedded through a parameterized nonlinear function, and then use the concept of inner product similarity by optimizing it with top-heavy ranking loss.

The main contribution of this paper is twofold. Firstly, a similarity function is proposed by optimizing the top-heavy ranking loss, which is actually designed for the re-id task. Secondly, use of a nonlinear embedding function is proposed to be used and to learn its inner product similarity with the structural learning framework.

II. METHODOLOGY

Here, we first introduce the nonlinear embedding function as proposed by us and then show the formulation of the similarity learning problem as an example case for the structured output learning problem. The various details are as under:

A. Nonlinear Embedding Function

We use a neural network to map the various metrics for person images of the actual feature space to the target feature space, and employ the inner product values as the similarity measures for the various values of person image pairs. Specifically, given a pair of person image features $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{d_1}$

$$s(\mathbf{x}_i, \mathbf{x}_j) =< f(\mathbf{x}_i), f(\mathbf{x}_j) > . \qquad (1)$$

Here $f(\mathbf{x})$ Is the nonlinear embedding function and is defined as:

the similarity between them is defined as:

$$f(\mathbf{x}) = \frac{\tanh(\mathbf{W}^T\mathbf{x} + \mathbf{b})}{\|\tanh(\mathbf{W}^T\mathbf{x} + \mathbf{b})\|_2}, \qquad (2)$$

where $\mathbf{W} \in \mathbb{R}^{d_1 \times d_2}$ Is a projection matrix $\mathbf{b} \in \mathbb{R}^{d_2}$ Is a bias Vector and d_1, d_2

Are dimensions of original feature space and

the target space respectively. $f(\mathbf{x})$ can be viewed as a single layer neural network followed by a L2-normalization layer. This normalization operation is important since it makes the similarity score to be irrelevant with magnitudes of embedded features. For the activation function, we have tested *relu*, *tanh* and *sigmoid*, and found *tanh* demonstrated better performance.

B. Similarity Learning with Structured Loss

Notations. The probe and gallery set are denoted $\mathcal{P} = \mathbf{x}^{p,u}$ as and $\{\mathbf{x}^{p,u}, id^{p,u}\}_{u=1}^{U}$ respectively. $\mathcal{G} = \{\mathbf{x}^{g,v}, id^{g,v}\}_{v=1}^{V}$ u $id^{p,u}$ U s the -th image, is its identity and is the total number of images in $\mathcal{P}$. Similar notations apply for variables of gallery set. Given a probe image $\mathbf{x}^{p,u}$, the gallery set is divided into its

relevant part and irrelevant part: $\mathcal{G} = \mathcal{R}^+(\mathbf{x}^{p,u}) \cup \mathcal{R}^-(\mathbf{x}^{p,u})$, where

$$\begin{aligned} \mathcal{R}^+(\mathbf{x}^{p,u}) &= \{\mathbf{x}^{g,v} | id^{g,v} = id^{p,u}, \mathbf{x}^{g,v} \in \mathcal{G}\}, \\ \mathcal{R}^-(\mathbf{x}^{p,u}) &= \{\mathbf{x}^{g,k} | id^{g,k} \neq id^{p,u}, \mathbf{x}^{g,k} \in \mathcal{G}\}. \end{aligned} \qquad (3)$$

In the following text, we will abuse the notation by simply write $\mathcal{R}^+$, $\mathcal{R}^-$ for $\mathcal{R}^+(\mathbf{x}^{p,u})$ and $\mathcal{R}^-(\mathbf{x}^{p,u})$ respectively without causing any confusion. Given the similarity function and a certain probe image, we expect its relevant images to be ranked before all irrelevant ones. Thus the output of the re-id task is a partial order:

$$\mathbf{y}^{p,u} = \{y^{p,u}_{v,k}\}, \quad y^{p,u}_{v,k} = \begin{cases} +1, & \mathbf{x}^{g,v} \succ \mathbf{x}^{g,k}, \\ -1, & \mathbf{x}^{g,v} \prec \mathbf{x}^{g,k}, \end{cases} \qquad (4)$$

where $\mathbf{x}^{g,v} \in \mathcal{R}^+$, $\mathbf{x}^{g,k} \in \mathcal{R}^-$, and $\mathbf{x}^{g,v} \succ \mathbf{x}^{g,k}(\mathbf{x}^{g,v} \prec \mathbf{x}^{g,k})$ represents that $\mathbf{x}^{g,v}$ is ranked before (after) $\mathbf{x}^{g,k}$. $\mathbf{y}^{p,u}$ is a set with total $N = |\mathcal{R}^+| \cdot |\mathcal{R}^-|$ elements.

Formulation as a Structural Learning Problem. We define a compatibility function that measures how well the output $\mathbf{y}^{p,u}$ matches the given input $\mathbf{x}^{p,u}$ as follows:

$$F(\mathbf{x}^{p,u}, \mathbf{y}^{p,u}; \mathbf{W}, \mathbf{b}) = \sum_{\substack{\mathbf{x}^{g,v} \in \mathcal{R}^+ }} \sum_{\substack{\mathbf{x}^{g,k} \in \mathcal{R}^-}} y^{p,u}_{v,k} \frac{s(\mathbf{x}^{p,u}, \mathbf{x}^{g,v}) - s(\mathbf{x}^{p,u}, \mathbf{x}^{g,k})}{N}.$$

III. EXPERIMENTS

A. Experimental Settings

To evaluate the effectiveness of the proposed SLTRL[1]method, we perform extensive experiments on VIPeR [29], The newly proposed LOMO [3] feature is used for all the following experiments and the dimensionality of LOMO feature is reduced to 600 by PCA. For our SLTRL method, we fix $d_2 = 200$ μ $d_1 = 600$ λ and . The learning rate and the 2×10^{-2} 10^{-2} regularization parameter are set as and respectively. These hyper-parameters are selected by using cross-validation. $\mathbf{b}$ $\mathbf{0}$ $\mathbf{W}$ The bias is initialized as and the weight matrix is initialized with 1 in diagonal and 0 otherwise. Both datasets are randomly divided into two subsets, one for training and the other for testing. Specifically, there are 316, the training sets for the VIPeR, dataset respectively. This partition is repeated 10 times to report the average result.

B. Experiments on VIPeR

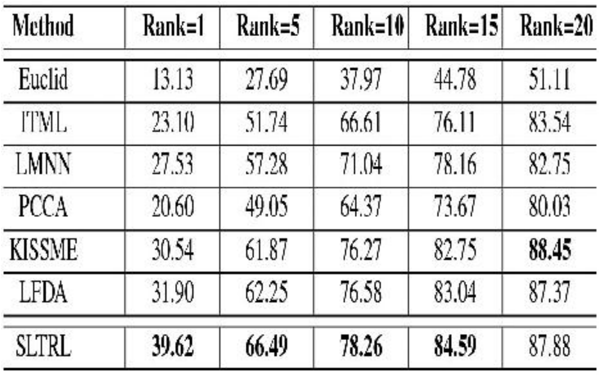
VIPeR [29] is a challenging person re-identification dataset which was widely used for performance evaluation. It contains 632 individuals captured in outdoor scenarios, and each person has two images observed from different camera views. All images are normalized to 128×48 for experiments.

Comparison with Metric Learning Algorithms. We first compare the proposed SLTRL algorithm with several conventional metric learning algorithms. Those methods include ITML [32], PCCA [13], LMNN [12], LFDA [20] and KISSME [14]. To make a fair comparison, we use the same LOMO feature and the same train/test split for each of the algorithms. The final results are listed in Table I. The CMC curves are shown in Fig. 2(a). It can be seen the SLTRL method performs better than the compared metric learning algorithms, especially at the top few ranks. Particularly, PCCA learns a linear transformation by optimizing the pairwise loss and LMNN optimizes the triplet loss. From the table we can see that our SLTRL method, which learns a nonlinear transformation and optimizes the topheavy listwise loss, consistently performs better than PCCA and LMNN.

Comparison with Different Losses. To verify the effectiveness of the proposed top-heavy ranking loss, we conduct experiments on VIPeR dataset with different loss functions.

TABLE I

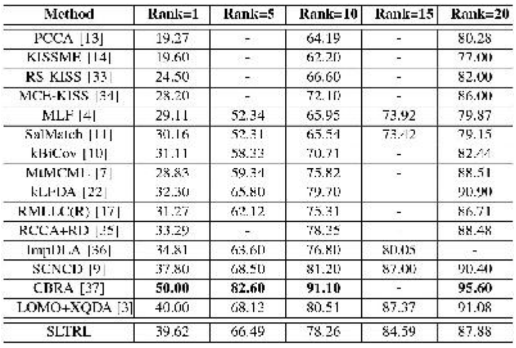
COMPARISON WITH DIFFERENT METRIC LEARNING ALGORITHMS WITH THE SAME FEATURE SET ON VIPER DATASET



Method	Rank=1	Rank=5	Rank=10	Rank=15	Rank=20
Euclid	13.13	27.69	37.97	44.78	51.11
ITML	23.10	51.74	66.61	76.11	83.54
LMNN	27.53	57.28	71.04	78.16	82.75
PCCA	20.60	49.05	64.37	73.67	80.03
KISSME	30.54	61.87	76.27	82.75	**88.45**
LFDA	31.90	62.25	76.58	83.04	87.37
SLTRL	**39.62**	**66.49**	**78.26**	**84.59**	87.88

TABLE II

COMPARISON WITH THE STATE-OF-THE-ART METHODS ON VIPER DATASET



Method	Rank=1	Rank=5	Rank=10	Rank=15	Rank=20
PCCA [13]	19.27	-	64.19	-	80.28
KISSME [14]	19.60	-	62.20	-	77.00
RS KISS [33]	24.50	-	66.60	-	82.00
MCH-KISS [34]	28.20	-	72.10	-	86.00
MLF [4]	29.11	52.34	65.95	73.92	79.87
SalMatch [11]	30.16	52.31	65.54	73.42	79.15
kBiCov [10]	31.11	58.33	70.71	-	82.44
MtMCML [7]	28.83	59.34	75.82	-	88.51
kLFDA [22]	32.30	65.80	79.70		90.90
RMLLC(R) [17]	31.27	62.12	75.31	-	86.71
RCCA+RD [35]	33.29	-	78.35		88.48
ImpDLA [36]	34.81	63.60	76.80	80.05	-
SCNCD [9]	37.80	68.50	81.20	87.00	90.40
CBRA [37]	**50.00**	**82.60**	**91.10**	-	**95.60**
LOMO+XQDA [3]	40.00	68.12	80.51	87.37	91.08
SLTRL	39.62	66.49	78.26	84.59	87.88

Δ_{AUC}, Δ_{TOP}, $\Delta_{0/1}$ And triplet loss(see Section III in the supplementary material). $\Delta_{0/1}$ Loss is set $a_1^N = 1$ And all a_j^N equal to 0. Fig. 2(b) shows the experimental results. It can be seen that among the four types of losses, learning with Δ_{TOP} achieves the best results which is in accordance with our expectation. $\Delta_{0/1}$ is the extreme case of the top-heavy loss with all weights concentrated on rank 1, however, it has the worst performance among the four loss functions possibly because it treats false ranks occurred from position 2 to N with no difference.

Comparison with the State of the Art. Finally, we compare the performance of the proposed SLTRL method with the state of-the-art results reported on the VIPeR dataset. The results are summarized in Table II. The CMC matching rates of ImpDLA are read from figures in [26]. From Table II we can see that the best performance is achieved by the CBRA [27] method, in which several complementary ranking lists are combined using ranking aggregation method. The proposed SLTRL method is complementary to CBRA because SLTRL can be used as a weak ranker to generate the original ranking list. Furthermore, our model is very flexible and can be easily integrated with feature learning methods like CNN [28] in a single framework to learn feature representation and similarity metric jointly.

IV. CONCLUSION

In this paper, we are able to propose an efficient nonlinear similarity learning method for person re-identification which is different from the already existing metric learning algorithms which can only optimize the pairwise, triplet or linear structured loss. We are able to optimize a top-heavy list wise loss also. Detailed experiments are conducted on VIPeR using the same set of features to show the better results with the adoption of the SLTRL method. Promising results are thus obtained on the challenging VIPeR datasets.

REFERENCES

[1] Y. Xie, H. Yu, X. Gong, Z. Dong, and Y. Gao, “Learning visual-spatial saliency for multiple-shot person re-identification,” *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1854–1858, Nov. 2015.

[2] G. Lisanti, I. Masi, A. Bagdanov, and A. Del Bimbo, “Person re-identification by iterative re-weighted sparse ranking,” *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1629–1642, Aug. 2015.

[3] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, “Person re-identification by local maximal occurrence representation and metric learning,” in *CVPR*, 2015, pp. 2197–2206.

[4] R. Zhao, W. Ouyang, and X. Wang, “Learning mid-level filters for person re-identification,” in *CVPR*, 2014, pp. 144–151.

[5] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, “Learning locally-adaptive decision functions for person verification,” in *CVPR*, 2013, pp. 3610–3617.

[6] W. Zheng, S. Gong, and T. Xiang, “Towards open-world person re-identification by one-shot group-based verification,” *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. PP, no. 99, pp. 1–1, 2015.

[7] L. Ma, X. Yang, and D. Tao, “Person re-identification over camera networks using multi-task distance metric learning,” *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3656–3670, 2014.

[8] X. Wang, W. Zheng, X. Li, and J. Zhang, “Cross-scenario transfer person re-identification,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, pp. 1–1, 2015.

[9] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li, “Salient color names for person re-identification,” in *Computer Vision–ECCV*, 2014, pp. 536–551.

[10] B. Ma, Y. Su, and F. Jurie, “Covariance descriptor based on bio-inspired features for person re-identification and face verification,” *Image Vis. Comput.*, vol. 32, no. 6, pp. 379–390, 2014.

[11] R. Zhao, W. Ouyang, and X. Wang, “Person re-identification by salience matching,” in *ICCV*, 2013, pp. 2528–2535.

[12] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, 2009.

[13] A. Mignon and F. Jurie, “Pcca: A new approach for distance learning from sparse pairwise constraints,” in *CVPR*, 2012, pp. 2666–2672.

[14] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in *CVPR*, 2012, pp. 2288–2295.

[15] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, “Relaxed pairwise learned metric for person re-identification,” in *Computer Vision–ECCV*, 2012, pp. 780–793.

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

[17] J. Chen, Z. Zhang, and Y. Wang, “Relevance metric learning for person re-identification by exploiting listwise similarities,” *IEEE Trans. Image Process.*, vol. PP, no. 99, pp. 1–1, 2015.

[18] X. Liu, H. Wang, Y. Wu, J. Yang, and M.-H. Yang, “An ensemble color model for human re-identification,” in *WACV*, 2015, pp. 868–875.

[19] J. Hu, J. Lu, and Y.-P. Tan, “Discriminative deep metric learning for face verification in the wild,” in *CVPR*, 2014, pp. 1875–1882.

[20] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, “Local fisher discriminant analysis for pedestrian re-identification,” in *CVPR*, 2013, pp. 3318–3325.

[21] K. Liu, Z. Zhao, and A. Cai, “Datum-adaptive local metric learning for person re-identification,” *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp.1457–1461, 2015.

[22] F. Xiong, M. Gou, O. Camps, and M. Sznaier, “Person re-identification using kernel-based metric learning methods,” in *Computer Vision–ECCV*, 2014, pp. 1–16.

[23] H. Liu, M. Qi, and J. Jiang, “Kernelized relaxed margin components analysis for person re-identification,” *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 910–914, Jul. 2015.

[24] D. Tao, L. Jin, Z. Yang, and X. Li, “Rank preserving sparse learning for kinect based scene classification,” *IEEE Trans. Cybernetics*, vol. 43, no. 5, pp. 1406–1417, 2013.

[25] D.Tao,J.Cheng,M.Song,andX.Lin,“Manifoldranking -basedmatrix factorization for saliency detection,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. PP, no. 99, pp. 1–1, 2015.

[26] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, “Large margin methods for structured and interdependent output variables,” J. Machine Learning Research 2005, pp. 1453–1484.

[27] N. Usunier, D. Buffoni, and P. Gallinari, “Ranking with ordered weighted pairwise classification,” in *Proc. 26th Annu. Int. Conf. Machine Learning*, 2009, pp. 1057–1064.

[28] Y. Yue, T. Finley, F. Radlinski, and T. Joachims, “A support vector method for optimizing average precision,” in *Proc. 30th Annu. Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, 2007, pp. 271–278.

[29] D. Gray, S. Brennan, and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” *IEEE Int. Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, vol. 3, no. 5, 2007.