PERSON RE-IDENTIFICATION VIA ADABOOST RANKING ENSEMBLE

Zhaoju Li *Zhenjun Han* *Qixiang Ye*

School of Electronics, Electrical and Communication Engineering
University of Chinese Academy of Sciences, Beijing, China
lizhaoju13@mails.ucas.ac.cn, {hanzhj, qxye}@ucas.ac.cn

ABSTRACT

Matching specific persons across scenes, known as person re-identification, is an important yet unsolved computer vision problem. Feature representation and metric learning are two fundamental factors in person re-identification. However, current person re-identification methods, which use single handcrafted feature with corresponding metric, could be not powerful enough when facing illumination, viewpoint and pose variations. Thus it inevitably produces suboptimal ranking lists. In this paper, we propose incorporating multiple features with metrics to build weak learners, and aggregate the base ranking lists by AdaBoost Ranking. Experiments on two commonly used datasets, VIPeR and CUHK01, show that our proposed approach greatly improves recognition rates over the state-of-the-art methods.

Index Terms— Person Re-identification, AdaBoost, Ranking, Base Model.

1. INTRODUCTION

Person re-identification targets at matching pedestrians across multiple cameras of non-overlapping fields-of-views (FOVs). With the increasing applications in security assurance and forensics, the research of person re-identification has been attracting increased attentions [1, 2]. Despite of the progress made in recent years, ample room for person re-identification research exists, particularly, in complex real world scenes of illumination, pose and viewpoint variations.

Feature representation and metric learning are two fundamental problems of person re-identification. For better representation, researchers have tried various features that are robust to appearance variations [3, 4, 5, 6]. Yang et al. [6] use a salient color names based color descriptor (SCNCD), which consists in a probability distribution over 16 color names. Kviatkovsky et al. [4] design illumination-invariant features which keep the relationship of colors in a log color space. Farenzena et al. [3] exploit the symmetry property of human body to build a robust description. On the other hand, metric learning [7, 8, 9] targets at maximizing intra-class distance and minimizing inter-class distance. For example, Zheng et al. [8] introduce a Probabilistic Relative Distance Comparison (PRDC) model to maximize likelihood of true matches. Li et al. [9] convert the person re-identification problem to multiple 0-1 classification problems, and propose the Locally-Adaptive Decision Function (LADF) to learn a locally adaptive threshold as classification criteria.

Considering that different features and metrics could deal with different challenges, feature/metric ensemble approaches including rank voting, probability ensemble [10] and SVM based ensemble [11, 12] have been investigated. Xiong et al. [10] propose a probability ensemble approach, which combines seven kinds of metrics on four kinds of features by multiplying their probabilities. Paisitkriangkrai et al. [12] train a structural SVM model to weight base metrics. Existing feature/metric ensemble approaches have reported performance gains; however, these ensemble approaches haven't fully explored the complementarity among features and metrics.

In this paper, we propose a new approach called AdaBoost ranking, which formalizes the ranking problem as a new 'classification' problem on permutations. Given features and metrics, we exhaustively match them to obtain base models. Base models play the roles of "weak classifiers", and the ensemble of them is a "strong classifier". The ensemble training procedure assigns weights to different base models using AdaBoost Ranking algorithm. It increases the weights of falsely matched samples. The two procedures of base model re-weighting and sample re-weighting alternate in each training iteration. After training, we first calculate the base ranking lists using learning based models, by which we calculate the final person re-identification results by a weighting scheme.

The contributions of the paper are as follows:

1. We propose AdaBoost ranking ensemble for person re-identification. Benefitting from the richer capacity in feature representation and metric learning, it can improve the diversity, coverage, and accuracy of models. Rather than extracting different features and concatenating them into a long vector, or putting empirical weights to metrics, our approach obtains the final model by a learning algorithm that optimizes the ranking results. Therefore, it is more adaptive to various video scenes. Negative feedback scheme puts more weights on the wrongly ranked samples. So samples being poorly ranked by one weak hypothesis could be compensated by an-

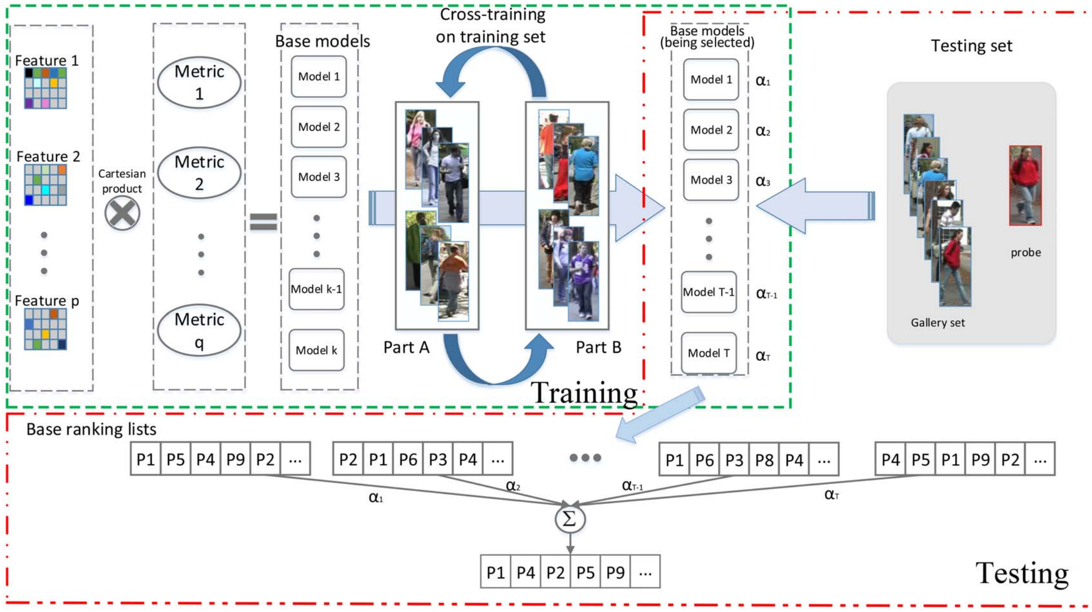

Fig. 1: The flowchart of the proposed approach. (Better Viewed in Color)

other.

2. We propose a cross-training strategy during the learning stage and adopt top-β threshold as the evaluation criteria of base models. These two strategies ensure the feasibility and reasonableness of AdaBoost ranking for person re-identification.

The rest of the paper is organized as follows. We describe the proposed approach in section 2, and present experimental results in section 3. We conclude the paper in section 4.

2. OUR APPROACH

Ranking is the key point of the person re-identification. That's because given a probe image and a gallery set, we need to rank the correct image before others, even not at 1st rank, but at least most of others. It is extremely hard to have a criterion to classify whether it is the corresponding image or not. What we do is to design a mechanism to make the correct one rank higher and make wrong images lower. All elaborated hand-crafted features and metrics are designed to achieve this goal. We'd like to refer these methods as weak models to person re-identification, as weak classifiers to AdaBoost. To this end, we adopt an AdaBoost framework to assemble weak models effectively and build final ranking list with "weak lists".

In this section, we detail the person re-identification approach demonstrated in Fig.1 by presenting the AdaBoost ranking method (Section 2.1) and base models (Section 2.2).

2.1. AdaBoost Ranking

We define sample image pairs $P_{i,g} = \{p_{i,j} = (x_i^p, x_j^g), j = 1, \dots, n\}$, where $p_{i,j}$ indicates the *i*-th person x_i^p from the probe set and x_j^g the *j*-th person from the gallery set. Therefore, $p_{i,j}$ represents a pair contains the same person from the two sets when $i = j$. $h(P_{i,g})$ is defined as a weak hypothesis, calculating base ranking list of the x_i^p with the whole gallery set based on each base model. $rank(x_j^g, h(P_{i,g}))$ defines the ranking position of the *j*-th person x_j^g in gallery set in base ranking list of $h(P_{i,g})$. $D_t(P_{i,g})$ defines the weight of x_i^p from the probe set, $D_t = \{D_t(P_{i,g}), i = 1, \dots, n\}$ means the distribution of images in the probe set in *t*-th iteration, which is updated in every loop. According to this distribution, we select the best weak hypothesis h_{k^*} in this iteration.

During the training stage, considering that current well-designed metric learning methods often overfit the training set, it is difficult to distinguish which one is the best hypothesis. To address this problem, we partition the training set into two parts: part A and part B, as shown in Fig. 1. One is used for training metric matrix M of base models, the other is used for evaluating such base models. we then alternate these two parts for training and evaluating in each iteration. Such cross-training strategy can avoid overfitting caused by training on the same set in every iteration. Since the state-of-the-art person re-identification usually performs under 50% accuracy for rank-1, we relax the constraint by forcing the corresponding image to be at top-β ($\beta = 5$ empirically) instead of at top-1.

Algorithm 1 AdaBoost Ranking

Given: $\boldsymbol{P} = [P_{1,g}, P_{2,g}, \ldots, P_{n,g}]$
Initialize $D_1(P_{i,g})$ (usually set to 1/n where n is the total number of samples).
for t=1, ..., T **do**
  choose $k^* = \operatorname{argmin}_k \epsilon_k$
    where $\epsilon_k = \sum_i D_t(P_{i,g}) f(h_k(P_{i,g}))$
  set base hypothesis $h_t = h_{k^*}$
  compute $\alpha_t = \frac{1}{2}\ln(\frac{1-\epsilon}{\epsilon})$
  update $D_{t+1}(P_{i,g}) = \frac{D_t(P_{i,g})\exp(-\alpha_t f(h_t(P_{i,g})))}{Z_t}$
    where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution)
end for
Output the final hypothesis $H(P) = \sum_t \alpha_t h_t(P)$

We formulate the classification of the weak hypothesis as

$$f(h(P_{i,g})) = \begin{cases} 1, & rank(x_i^g, h(P_{i,g})) \leq \beta \\ -1, & otherwise. \end{cases} \tag{1}$$

Following the procedure of Algorithm 1, each selected base model has a ranking list in the testing stage. We aggregate the final rank by these ranking lists as

$$H(P) = \sum_t \alpha_t h_t(P), \tag{2}$$

where P denotes the samples of probe set, and α_t denotes the weight of weak hypotheses h_t at the t-th iteration, as detailed in Algorithm 1. And $h_t = h_{k^*}$ is selected based on the minimal error rate ϵ_k of all base models. In a word, we obtain the final ranking list by adaptively assembling base ranking lists.

2.2. Base Models

We obtain base models by exhaustively combining the features and metrics.

2.2.1. *Features*

Based on the assumption that richer features can better represent the persons, we utilize four kinds of off-the-shelf features to represent the persons in our approach.(Some researchers adopt single type of features [13, 14]. We don't follow them because those one type feature descriptors are too weak to construct a weak learner, even loosening the rank position to top-β.) The features include:

1.Background weighted Color naming (CN) and HSV histogram. CN [15] project color space into 11 color names. We assign an empirical weight to integrate foreground and background information to extract HSV and CN histogram.

2. Local Maximal Occurrence Feature (LOMO). LOMO [16] uses multiscale Retinex and Scale Invariant Local Ternary Pattern (SILTP), which contains vivid color information and illumination invariant texture.

3. Ensemble of Localized Features (ELF6). ELF6 [8] is computed from histograms in six equally divided horizontal stripes. 8 color channels and 21 texture filters are used for histogram representation.

4. Weighted Histograms of Overlapping Stripes (WHOS). WHOS [17] is based on coarse, striped pooling of local features.

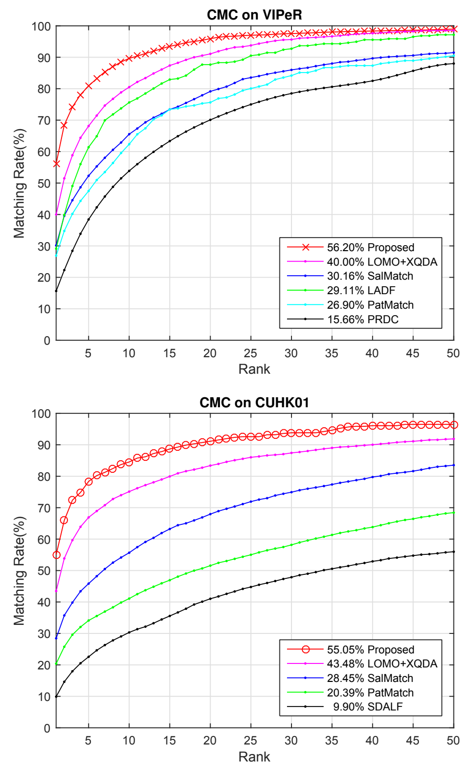

Fig. 2: CMC comparison on VIPeR and CUHK01 datasets.

2.2.2. *Metrics*

Based on the feature representation above, we select three effective methods to perform metric learning.

1. Kernel Local Fisher Discriminant Classifier (kLFDA) [10]. It is a nonlinear extension of LFDA [18], which conducts supervised dimensionality reduction on LFDA.

2. Locally-Adaptive Decision Function (LADF) [9]. LADF learns a locally adaptive threshold as classification criteria.

3. Cross-view Quadratic Discriminant Analysis (XQDA) [16]. It is an extension of Bayesian face and KISSME [7] approaches to cross-view metric learning.

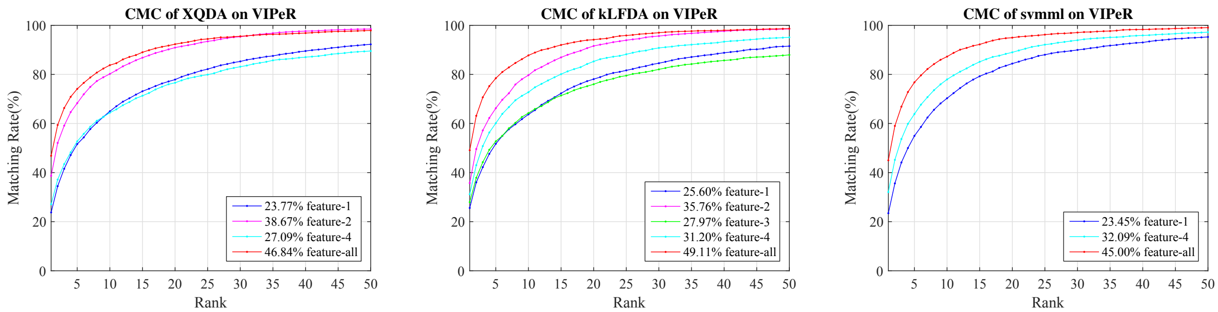

Fig. 3: Evaluation of feature ensembles.

3. EXPERIMENTS

We evaluate our proposed approach on two commonly used public datasets: VIPeR and CUHK01. To quantitatively evaluate the effectiveness of the proposed approach, the cumulative match characteristic (CMC) [19] curve is used for performance evaluation. We ran 10 trials and reported the average performance of experimental results.

3.1. Datasets

VIPeR [14] is a challenging and popular dataset for person re-identification. There are 632 pedestrians with resolution 128×48. Each person has two images from different cameras. In our experiments, we randomly select half of the image pairs (i.e. 316 pairs) for training , and the rest for testing.

CUHK01 [20] is another widely used dataset. There are 971 pedestrians with resolution 160×60. Each person has four images from different views. We adopt single shot experiment setting, to randomly select one image from each view, and then randomly select half of the image pairs (486 pairs) for training, and the rest for testing.

Table 1: Top rank comparisons with the state-of-the-arts on the VIPeR and CUHK01 datasets.

Method	VIPeR(p=316)				
	r=1	r=5	r=10	r=20	r=50
PRDC [8]	15.66	38.42	53.86	70.09	80.00
LADF [9]	29.11	61.39	75.63	87.66	97.15
PatMatch [5]	26.90	47.47	62.34	75.63	90.51
SalMatch [21]	30.16	52.31	65.54	79.15	91.49
Ensb [10]	36.1	68.7	80.1	85.6	-
LOMO+XQDA [16]	40.00	64.40	80.51	91.08	98.54
CMC^{top} [12]	45.89	77.5	88.9	95.8	-
Proposed	**56.20**	**80.92**	**89.65**	**95.92**	**98.96**

Method	CUHK01(p=486)				
	r=1	r=5	r=10	r=20	r=50
SDALF [3]	9.90	22.57	30.33	41.03	55.99
PatMatch [5]	20.39	34.12	41.09	51.56	68.42
SalMatch [21]	28.45	45.85	55.68	67.95	83.53
LOMO+XQDA [16]	43.48	66.91	75.11	83.38	91.88
CMC^{top} [12]	53.68	76.11	84.21	90.53	-
Proposed	**55.05**	**78.35**	**84.53**	**91.13**	**96.49**

3.2. Performance and comparison

In Fig. 2 and Table 1, we compare our approach with the recent representative approaches [8, 9, 5, 21, 10, 12, 16]. Among these methods, [12] is the structural SVM learned ensemble approach, and [10] is the probability ensemble approach. The rest of them are simplex. As shown in Table 1, on the VIPeR dataset, our approach results in consistently better performance than the other approaches. Specifically, the rank-1 accuracy of our approach (56.20%) is 10% higher than [12] and 16% higher than [10]. Our approach also achieves the state-of-the-art results (55.05%) on CUHK01.

For further validation, we fix single metric and compare the results of single base models with the ensemble of all features by our method on VIPeR. In Fig. 3, it can be seen that LOMO+XQDA is the best base model (38.67%) which is proposed in [16]. Our final result is over 17% higher than it. Three subfigures show that using only feature ensembles also gains performance improvement from base models through our approach (8.2%, 13.4%, 13.0% respectively). Our final improvement of recognition accuracy are two stages: the first stage is the ensemble of features, the second stage is the ensemble of the results of above stage.

4. CONCLUSION

We have proposed a simple but effective AdaBoost ranking ensemble approach for person re-identification. Experimental results with comparisons to other representative methods are provided, which indicate that the proposed approach outperforms other ensemble based person re-identification approaches, and achieves state-of-the-art performance.

5. ACKNOWLEDGEMENT

This work is supported in part by the National Science Foundation of China under Grants 61271433, 61202323 and 61401426, and Beijing Municipal Science & Technology Commission.

6. REFERENCES

[1] Apurva Bedagkar-Gala and Shishir K Shah, "A survey of approaches and trends in person re-identification," *Image and Vision Computing*, vol. 32, no. 4, pp. 270–286, 2014.

[2] Gianfranco Doretto, Thomas Sebastian, Peter Tu, and Jens Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 2, pp. 127–151, 2011.

[3] Loris Bazzani, Marco Cristani, and Vittorio Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 130–144, 2013.

[4] Igor Kviatkovsky, Amit Adam, and Ehud Rivlin, "Color invariants for person reidentification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 7, pp. 1622–1634, 2013.

[5] Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Unsupervised salience learning for person re-identification," in *CVPR*, 2013, pp. 3586–3593.

[6] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z Li, "Salient color names for person re-identification," in *ECCV*, pp. 536–551. 2014.

[7] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *CVPR*, 2012, pp. 2288–2295.

[8] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Reidentification by relative distance comparison," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 3, pp. 653–668, 2013.

[9] Zhen Li, Shiyu Chang, Feng Liang, Thomas S Huang, Liangliang Cao, and John R Smith, "Learning locally-adaptive decision functions for person verification," in *CVPR*, 2013, pp. 3610–3617.

[10] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier, "Person re-identification using kernel-based metric learning methods," in *ECCV*, pp. 1–16. 2014.

[11] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, Tao Xiang, and Q Mary, "Person re-identification by support vector ranking.," in *BMVC*, 2010.

[12] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *CVPR*, 2015.

[13] Cheng-Hao Kuo, Shamsul Khamis, and Vinay Shet, "Person re-identification using semantic color names and rankboost," in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pp. 281–287.

[14] Douglas Gray and Hai Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *ECCV*, pp. 262–275. 2008.

[15] Joost Van De Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus, "Learning color names for real-world applications," *Image Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1512–1523, 2009.

[16] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li, "Person re-identification by local maximal occurrence representation and metric learning," in *CVPR*, 2015, pp. 2197–2206.

[17] G. Lisanti, I. Masi, A.D. Bagdanov, and A. Del Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *PAMI, IEEE Transactions on*, vol. 37, no. 8, pp. 1629–1642, 2015.

[18] Sateesh Pedagadi, James Orwell, Sergio Velastin, and Boghos Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *CVPR*, 2013, pp. 3318–3325.

[19] Douglas Gray, Shane Brennan, and Hai Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007, vol. 3.

[20] Wei Li, Rui Zhao, and Xiaogang Wang, "Human reidentification with transferred metric learning.," in *ACCV (1)*, 2012, pp. 31–44.

[21] Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Person re-identification by salience matching," in *ICCV*, 2013, pp. 2528–2535.