Optimizing Mean Reciprocal Rank for Person Re-identification

Yang Wu, Masayuki Mukunoki, Takuya Funatomi, Michihiko Minoh
Academic Center for Computing and Media Studies, Kyoto University, Kyoto 606-8501, JAPAN
{yangwu,mukunoki,funatomi,minoh}@mm.media.kyoto-u.ac.jp

Shihong Lao
Omron Corporation, Kyoto 619-0283, Japan
lao@ari.ncl.omron.co.jp

Abstract

Person re-identification is one of the most challenging issues in network-based surveillance. The difficulties mainly come from the great appearance variations induced by illumination, camera view and body pose changes. Maybe influenced by the research on face recognition and general object recognition, this problem is habitually treated as a verification or classification problem, and much effort has been put on optimizing standard recognition criteria. However, we found that in practical applications the users usually have different expectations. For example, in a real surveillance system, we may expect that a visual user interface can show us the relevant images in the first few (e.g. 20) candidates, but not necessarily before all the irrelevant ones. In other words, there is no problem to leave the final judgement to the users. Based on such an observation, this paper treats the re-identification problem as a ranking problem and directly optimizes a listwise ranking function named Mean Reciprocal Rank (MRR), which is considered by us to be able to generate results closest to human expectations. Using a maximum-margin based structured learning model, we are able to show improved re-identification results on widely-used benchmark datasets.

1. Introduction

The widely deployed monitoring cameras in both restricted buildings and public areas can easily been organized into a camera network for surveillance, healthcare, entertainment and business applications. However, the development of reliable algorithms for such applications are far behind that of the hardware and networking. One of the most challenging but unsolved problems is person re-identification, *i.e.*, identifying an human individual again upon re-entering the scene after some time [6], and that person may be across cameras. The difficulties include, but are not limited to viewpoint, pose and illumination changes.

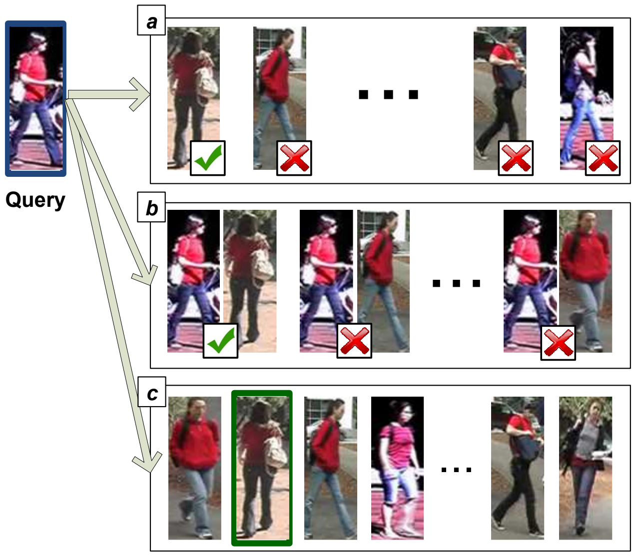

Figure 1. Motivation of the work. Existing approaches on person re-identification belong to either group (a) binary yes/no verification or group (b) pairwise relevant/irrelevant classification. These two groups both aim at directly outputting the correct match. Differently, we propose a third way (c): optimizing a listwise loss function named MRR (Mean Reciprocal Rank) to bias on placing the correct match at the front of the rank list, which coincides with the demands in practical video surveillance applications.

An ideal person re-identification system should be able to find all the images containing the same person from a large gallery without any false alarms. However, in practice we can tolerate some imperfectness when working with friendly user interfaces. For example, in the surveillance scenario, usually the user would like to find the correct match by himself/herself from a few candidates listed on the screen of a search/retrieval system. In other words, it is not expected that the system can automatically rank the correct matches in front of the incorrect ones, and it is acceptable if it can place these correct ones in the first few candidates (better to fit in a screen). Actually, the most widely used evaluation criterion for person re-identification is the CMC (Cumulative Matching Characteristic) curve, which to some extent coincides with such a practical expectation. This criterion shows how performance (recognition/reacquisation rate) improves as the number of requested images increases, and CMC(k) indicates the rank k recognition rate.

Unfortunately, the CMC curve is not a single-trial measurement, so it cannot be directly optimized. Despite that, there are some other criteria which may potentially coincide with the CMC curve to some extent. MRR (Mean Reciprocal Rank) seems to be the closest one among all the existing candidates. The reciprocal rank of a query response is the multiplicative inverse of the rank of the first correct match, and MRR is the average of such reciprocal ranks of results over the whole query set [14]. Such a criterion encourages but not strongly forces the correct matches to be ranked before all the incorrect ones. However, we find that there is no existing work on optimizing MRR or other similar criteria for person re-identification, and all the former learning-based efforts are either on pointwise binary yes/no verification [7] or pairwise relevant/irrelevant classification [2, 12]. Instead, this paper aims at directly optimizing MRR for person re-identification by exploring a maximum-margin based structured learning framework, as shown in Figure 1. It results in significantly better performance than other approaches on commonly used benchmark datasets.

2. Related work

Traditionally, person re-identification is treated as a recognition problem. Therefore, a straight-forward approach is to utilize powerful learning algorithms like AdaBoost [7] to learn a model from weak features. Unfortunately, the performance is limited due to the great difficulties of the problem itself (large within-class appearance variations and small between-class variations). Instead of relying on strong classifiers, there are also some efforts on exploring features for similarity/distance based matching. Along with the publication of the first benchmark dataset on viewpoint invariant pedestrian recognition (VIPeR), some simple features together with common distance metrics were tested as baseline algorithms [6]. Later on, a metric learning algorithm named LMNN (Large Margin Nearest Neighbor) was explored and extended for learning a discriminative matching model [2]. Feature representation has also been greatly improved recently by new hand-designed structures like the silhouette partition axes based on symmetry and asymmetry of the human body [5], leading to boosted performance without supervised learning.

There was no effort on treating person re-identification as a ranking problem until very recently[12]. The motivation of this work is to learn a subspace where the potential true match (relevant query-answer pair) can get higher rank than any direct distance measure. An Ensemble RankSVM algorithm was proposed to combine a set of weak RankSVM rankers into a stronger one, which avoids the selection of the regularization parameter C while at the same time is computationally more efficient. However, this approach is still limited to relevant/irrelevant pairwise classification, but not listwise ranking. The strong partial order constraints are theoretically promising but hard to satisfy when dealing with real data. Therefore, the performance is limited similar to traditional classification-based methods.

Actually, listwise learning to rank algorithms have been explored and widely used in some other research fields including document retrieval, text-based search engine, image similarity learning and image retrieval. A good example is the research on maximum-margin based structured learning algorithms, which is now able to optimize many non-smooth listwise ranking losses like MRR [1]. Metric learning techniques have also been greatly explored, and some of them work together with maximum-margin based structured learning techniques [11]. In this paper we will explore such accomplishments to optimizing MRR for better person re-identification performances.

The rest of this paper is organized as follows. Section 3 presents a generic maximum-margin based model and our discussions on loss function, feature map, optimization and inference within this model. Experimental results on two representative benchmark datasets are shown in section 4, while conclusions are given in section 5.

3. Structured learning for ranking

Given a set of queries $\mathcal{Q}$ and a corpus/gallery $\mathcal{X}$, the desired ranking model should be able to properly rank all the items in $\mathcal{X}$ for each query $q \in \mathcal{Q}$. In real applications like person re-identification, usually the absolute position of each item $x \in \mathcal{X}$ [1] in the rank is not very important, and we are more interested in grouping the corpus/gallery into two subsets according to their relationships with the query q: one is composed by relevant items while the other has only irrelevant ones in it. We will use $\mathcal{X}_q^+$ and $\mathcal{X}_q^-$ to denote these two subsets. Let $\mathcal{Y}$ denote the set of all possible permutations (rankings) of $\mathcal{X}$, and $x_{qi} \succ x_{qj}$ indicate that for the ith and jth items $x_{qi}, x_{qj} \in \mathcal{X}$ w.r.t. q, x_{qi} is ranked higher than x_{qj}. In the training stage, $\mathcal{X}_q^+$ and $\mathcal{X}_q^-$ are given and the ranking model is expected to place the whole set of $\mathcal{X}_q^+$ before that of $\mathcal{X}_q^-$, while the orders within these two subsets do not matter. During testing, $\mathcal{X}_q^+$ and $\mathcal{X}_q^-$ are unknown and the whole set of $\mathcal{X}$ will be ranked, expecting that similarly relevant items are placed before irrelevant ones.

[1] Usually data items are represented by feature vectors and mathematically they should be denoted by bolded characters, however in this paper we use normal ones instead for simplicity if not specifically defined.

Recently, structured learning has become an active research topic in machine learning for predicting structured outputs, which is a common demand in many real applications. Since ranking is a typical structured output prediction problem whose desired outputs are proper permutations/ordering of items, structured learning has been greatly explored for it [9]. Inspired from the literature, we present a generic structured learning model based on maximum-margin formulation and discuss the key components of it along with its optimization and inference methods.

3.1. A generic maximum-margin based model

Suppose $\mathcal{X}, \mathcal{Q} \subset \mathbb{R}^d$ and $y \in \mathcal{Y}$ is a ranking of $\mathcal{X}$ w.r.t. q. We use $\phi_{qi} = \phi(x_{qi}, q)$ to denote the relative feature representation of x_{qi} w.r.t. q, then a desired ranking model can be defined as a linear function $f_w(x_{qi}) = w^T \phi_{qi}$ for scoring x_{qi}, where $w \in \mathbb{R}^d$, and the ranking is done by sorting the score in a descending order. To learn the weight vector w of the model, usually a vector-valued joint feature map $\psi(q, y, \mathcal{X}) \in \mathbb{R}^d$ is adopted to represent the whole set of ranked data $\mathcal{X}_q$, and the best w is the one that simultaneously makes $y_q^* = \arg\max_y w^T \psi(q, y, \mathcal{X}), \forall y \in \mathcal{Y}$ for all $q \in \mathcal{Q}$, where y_q^* is the ground truth ranking of $\mathcal{X}$ for q.

To differentiate y_q^* from the other ys, we should define a loss function denoted by $\Delta(y_q^*, y)$ which penalizes predicting y instead of y_q^* by forcing $\Delta(y_q^*, y) \geq 0, \forall y \in \mathcal{Y}$ with $\Delta(y_q^*, y_q^*) = 0$ and usually larger values when $y \neq y_q^*$. Then the learning of w can be achieved by solving the following maximum-margin based optimization problem:

$$\arg\min_w \frac{1}{2}\|w\|^2 + \frac{C}{|\mathcal{Q}|}\sum_q \xi_q \tag{1}$$

$$s.t. \quad \begin{aligned} w^T\psi\left(q, y_q^*, \mathcal{X}\right) \geq w^T\psi\left(q, y, \mathcal{X}\right) + \Delta\left(y_q^*, y\right) - \xi_q, \\ \forall q, \forall y \neq y_q^*; \\ \xi_q \geq 0, \forall q \in \mathcal{Q}. \end{aligned}$$

where ξ_q is the slack variable for q and C is the trade-off parameter for balancing the margin and the training error.

During testing, given an input query q, the model will return the ranking of $\mathcal{X}$ (*i.e.* y_p) for it:

$$y_q = \arg\max_{y \in \mathcal{Y}} w^T\psi\left(q, y, \mathcal{X}\right). \tag{2}$$

In practice, this can be done by simply sorting the score $f_w(x_{qi}) = w^T\phi_{qi}$ generated from each $x_{qi} \in \mathcal{X}$ in a descending order.

3.2. Loss function and feature map

In the model presented above, there are two critical components whose definitions greatly influence the properties and performances of it: the loss function $\Delta(y_q^*, y)$ and the feature map $\psi(q, y, \mathcal{X})$. However, they are domain-specific, and the proper collaboration between them hasn't been well understood [1]. Nevertheless, we discuss here how they could be designed for the person re-identification problem.

For ranking, ideally we expect the model can correctly order each relevant-irrelevant pair of gallery items, *i.e.* $f_w(x_{qi}^+) > f_w(x_{qj}^-), \forall x_{qi}^+ \in \mathcal{X}_q^+, x_{qj}^- \in \mathcal{X}_q^-$. However, this is usually unrealistic in real applications, and during the process of model learning such an ideal partial ordering doesn't hold due to the imperfect w being optimized. Therefore, we need some loss function to quantitatively evaluate how far we are away from the ideal (*i.e.* performance evaluation), and such a loss function can also guide the optimization of the model w. Intuitively, the loss function represents our expectations about the capabilities of the model. As presented in [10], for information retrieval which is one of the typical applications of ranking, publicly available loss functions or evaluation criteria can be grouped into three types:

1. **Pointwise** functions. These functions treat the query-candidate pairs independently in a regression manner (*e.g.* $\sum_{i=1}^{|\mathcal{X}|} |f_w(x_{qi}) - z_{qi}|$ where z_{qi} is the binary relevance indicator, *i.e.* $z_{qi} = 1$ means x_{qi} is relevant to q and $z_{qi} = 0$ is the opposite).

2. **Pairwise** functions. In such functions, the average number of inversions in ranking is minimized (*e.g.* $\sum_{i=1}^{|\mathcal{X}_q^+|}\sum_{j=1}^{|\mathcal{X}_q^-|} [\![f_w(x_{qi}^+) < f_w(x_{qj}^-)]\!]$ where $[\![\cdot]\!]$ denotes the Iverson bracket [8]), which usually done by learning a binary classifier for the pairwise partial ordering.

3. **Listwise** functions. These functions are undecomposable statistics based on the whole query set and thus they are often non-smooth and uneasy to optimize directly. Examples are AUC (area under ROC curve), Prec@k (Precision-at-k), MAP (mean average precision), MRR (mean reciprocal rank), NDCG (normalized discounted cumulative gain), *etc.* [10, 11]

As it has been stated before, the loss functions used for person re-identification are either pointwise [7] or pairwise [12], however, the dominating evaluation criterion for such a task is the CMC curve which is listwise. Such a mismatch may have limited the performance of existing approaches whether they are purely matching/classification based or ranking-based. The major goal of this paper is to explore listwise loss functions for a relatively more direct optimization of the listwise evaluation criterion.

The CMC curve shows how performance (recognition/reacquisation rate) improves as the number of requested images increases. It was first introduced for evaluating person re-identification performance in [6] , thought to be a better evaluation method than traditional methods which were designed for detection instead of ranking. However, such a ranking-based criterion hasn't been seriously investigated for designing a proper loss function for ranking. Besides of that, we argue that the CMC curve is not a well-suited criterion for person re-identification though it is informative. One shortage of it is that it may misleading people to focus on the first few points (top k ranked items) without concerning the size of the gallery/corpus. Such a problem has been overcome in [6] by introducing another criterion named synthetic disambiguation/reacquisition rate (SDR/SRR) derived from CMC. Another shortage is that it is better to be treated as whole curve but not a single condensed value, thus it's hard to transfer it into a loss function for learning. In [6], the authors used the area under the CMC curve for cross-validation, but it will confuse the bonuses got from boosting the ranks of relevant items at different positions. More concretely, when a relevant item is boosted from rank k to rank $k-l$, the gain of the area under the CMC curve will be the same when k varies as long as l is fixed. Therefore, we have to explore other loss functions (coinciding with the CMC curve) for model learning. However, we think that the CMC curve and SDR/SRR curve are still good for performance evaluation. If the performance is not compared across datasets with different sizes, using the CMC curve itself is enough.

Among all publicly available losses, we believe that MRR is the most proper one for person re-identification. Similar to the CMC curve, it also only focuses on the first relevant item in the rank y, which coincides with the user expectation for a person re-identification system. Formally, it can be expressed as:

$$S_{mrr}\left(q,y\right)=\frac{1}{|\mathcal{Q}|}\sum\nolimits_{q\in\mathcal{Q}}\left\{\begin{array}{ll}1/r_q, & r_q<k\\ 0, & r_q\geq k\end{array}\right. \quad (3)$$

where r_q is the position of the first relevant item w.r.t. q in y, and k is the number of top ranked items to be considered. Unlike the CMC curve, MRR is a single condense value which is sensitive to the absolute positions of the first relevant item for each query. Therefore, it is a good candidate for our desired loss function, and also an good alternative to the CMC curve for performance evaluation. The loss function based on MRR can be as simple as:

$$\Delta_{mrr}\left(y_q^*,y\right)=1-S_{mrr}\left(q,y\right), \quad (4)$$

since for the ground truth y_q^*, $S_{mrr}\left(q,y_q^*\right)=1$.

Performance evaluation criteria directly determine the experimental results for algorithm/system comparison, while loss functions determine the properties of learned ranking models. However, it has been demonstrated that the feature map in the structured learning model also strongly influences the training process and results [1]. Though it is hard to design a well-suited feature map for a given loss function, the feature map $\psi_{mrr}(q,y,\mathcal{X})$ proposed in [1] is the most promising one we have found for the MRR loss function, therefore we decided to use it for our structured learning model. We omit the details of its derivation for brevity (readers are referred to [1] if interested), and just present the definition of it:

$$\psi_{mrr}\left(q,y,\mathcal{X}\right)=\sum\nolimits_{j:j\succ i_1(y)}\left(x_{qj}^{-}-x_{qi_1(y)}^{+}\right) \quad (5)$$

where $i_1(y)$ denotes the rank of the top-ranking relevant item $x_{qi_1(y)}^{+}\in\mathcal{X}_q^{+}$. Though very simple, such a feature map still ensures that finding the best $y_q=\arg\max_y w^T\psi_{mrr}(q,y,\mathcal{X})$ is equivalent to sorting the gallery items by descending their scores generated by $f_w(x_{qi})=w^T\phi_{qi}$, in which $\phi_{qi}=q-x_{qi}$.

3.3. Optimization and inference

As shown in equation 1, the optimization of the model is to minimize the objective function with respect to the partial ordering constraints whose number is exponential w.r.t. $|\mathcal{X}|$, therefore an exact optimization by taking into account of all the constraints is usually computationally infeasible. A common strategy is to perform an approximate optimization using a cutting-plane algorithm, which maintains a working set of constraints by finding the most violated constraint at each step based on the current model w and slacks and then update the model and slacks based on the new constraint set. Such an iteration goes on until the most violated partial ordering constraint is within some predefined tolerance ϵ. Theoretically, it can be proved that for any fixed ϵ, a constant number of violators will be explored before convergence. Therefore, the overall training process of the model can terminate in linear-time. Details of the algorithm can be found in [9].

In the optimization process of the cutting-plane algorithm, there is an inference problem embedded in each iteration: finding the most violated constraint. Though w, ξ_qs and ϵ are all fixed, without special treatment, the search space of y is still an exponential number w.r.t. $|\mathcal{X}|$. Fortunately, the properties of the loss function can be utilized to greatly reduce the search space. An efficient inference algorithm has been proposed in [1] for optimizing the structured SVM model with the MRR loss, and thus the model is named "SVMMRR", similar to "SVMMAP"[15] which optimizes the MAP loss.

Though SVMMRR is a straight-forward solution to the structured learning model with MRR loss, its performance is not as good as SVMMAPon standard text retrieval benchmark datasets [1]. However, recently a new model named MLR (Metric Learning to Rank) with a couple of changes to the maximum-margin based structured learning model shows significant performance and efficiency improvements even when the same loss function like MAP is used for training [11]. The two major changes are: constraining the model parameters to be an symmetric PSD (positive semi-definite) matrix W instead of a unconstrained vector w, and replacing the l_2 regularizer $\frac{1}{2}\|w\|^2$ with a new one in the l_1 form: $tr(W)$, forcing W to be sparse. And according, a slightly different optimization algorithm is proposed. As pointed out by the authors, the symmetry and the PSD constraint of W may have limited the modeling capability of MLR, however, it's hard to say W is less powerful than w because W has a higher dimensionality. It is also unclear that whether the change of the objective function has improved the model or not. A theoretical analysis is beyond the scope of this paper, so we just choose MLR as the instantiation of the structured learning model to demonstrate the superiority of optimizing MRR for person re-identification.

4. Experimental results

Feature Representation: Recently, a lot of work has been done on feature representation for person re-identification, especially on the appearance-based features [3]. Among all these efforts, the SDALF (symmetry-driven accumulation of local features)[5] may be the most carefully hand-designed feature set. Though worked with very simple matching models without learning, it has shown significant superiority against other state-of-the-art algorithms (some are using powerful classifiers) via extensive experiments. However, SDALF cannot be directly used in our model because two of its three types of features have variable size and dimensionality, which are ok or even beneficial for matching but unsuitable for learning-based classification and ranking. Therefore, we only use the weighted color histograms in our experiments. Such a group of features are based on background subtraction and carefully designed middle axes extraction. Though it is selective and robust, it may be too restricted even in the view of color features. Therefore, we use the densely sampled color histograms proposed in [2] as a supplementary descriptor.

Datasets: We test our model on the two publicly available and widely used benchmark datasets: VIPeR dataset [6] and ETHZ dataset [4, 5, 13]. VIPeR dataset contains 632 pedestrian image pairs taken by two cameras from different viewpoints (no less than 45° view angle changes) at different outdoor places, thus the illumination, body pose and the background in the image pair of the same person may be also different. All the pedestrians are cropped and scaled to 128×48 pixels. This dataset is most widely used and regarded to be the most challenging public dataset for pedestrian re-identification [5]. The ETZH dataset contains three video sequences of crowed street scenes captured by two moving cameras mounted on a chariot, and it was originally used for pedestrian detection. We use the three subsets of it which were extracted by Schwartz and Davis for person re-identification [13]. Compared to VIPeR, it has smaller pose and viewpoint variations, but more occlusions. We normalized all the images to 64×32 pixels as in [5]. The details of the 3 sequences are: 83 pedestrians from 4,857 images in SEQ. #1, 35 pedestrians from 1,936 images in SEQ. #2, and 28 pedestrians within 1,762 images in SEQ. #3. Like the single-shot case in [5], we split each dataset into equally sized training and test sets, in which only one image was sampled for each person to form the corpus, and all the other images served as queries. The splitting and sampling were done randomly and repeated 10 times for averaging. For a fair comparison to the original SDALF results, we used the experimental data provided by the authors of [5].

Methods for Comparison: We compare our model (named "OMRR") to two other representative methods which have the state-of-the-art performance on person re-identification. One is the matching-based model using SDALF feature set [5], and the other is the metric learning model based on Large Margin Nearest Neighbor (LMNN) approach [2]. Note that in this paper we used the original model of LMNN because it performs almost the same as the improved model proposed in [2]. Since the weighted HSV color histogram features (denoted by "wHSV" here) are from SDALF while the densely sampled color histograms (using "DCHs" for short) are from [2], we compare our model with both of them using the same feature sets: wHSV, DCHs, and wHSV-DCHs (wHSV and DCHs together) for a fair comparison. Besides that, we also show the results of the original SDALF model (using color, region and texture features) on VIPeR dataset for a clear reference.

Results: Experimental results are evaluated by both the CMC curves (see Figure 2 and Figure 3) and the MRR values (see Table 1). As it can be seen, generally speaking, the proposed model generates significantly better ranking results than other methods when the same features are used. An obvious advantage of our model is that it tends to put the correct matchings in the first few candidates of the rank, therefore resulting in a significantly higher MRR.

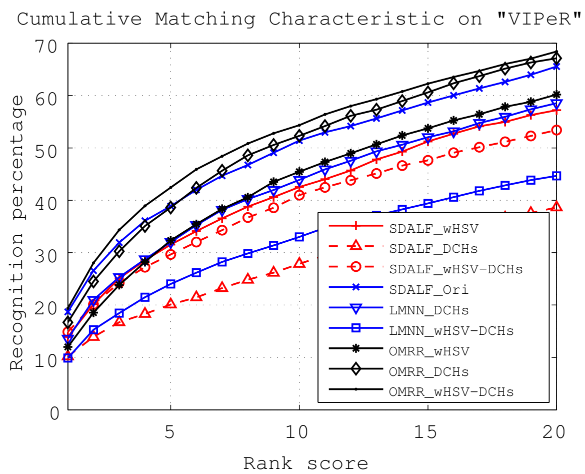

Figure 2. CMC performance comparison on VIPeR.

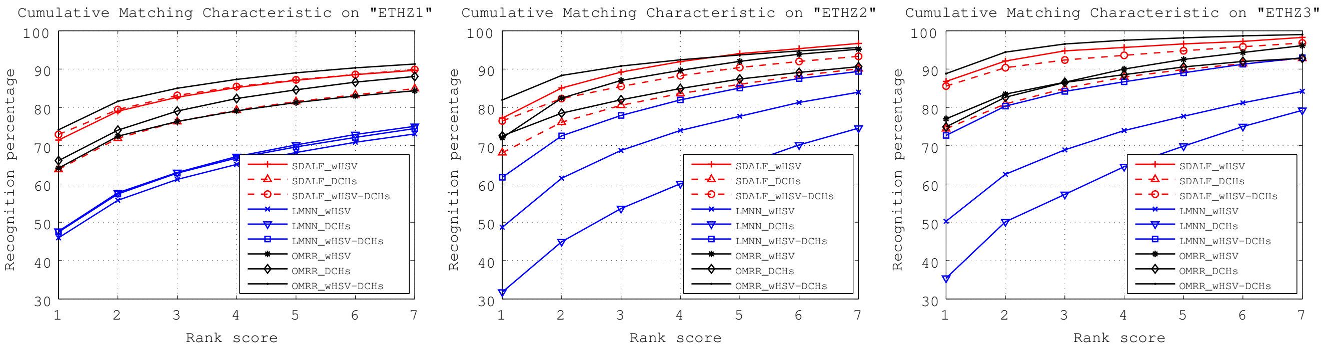

Figure 3. CMC performance comparison on ETHZ1, ETHZ2, and ETHZ3.

Method	VIPeR	ETHZ1	ETHZ2	ETHZ3
SDALF_wHSV	0.24	0.79	0.84	0.91
SDALF_DCHs	0.16	0.72	0.76	0.81
SDALF_wHSV-DCHs	0.23	0.79	0.83	0.90
SDALF_Ori [5]	0.29	-	-	-
LMNN_wHSV	-	0.57	0.62	0.63
LMNN_DCHs	0.24	0.58	0.48	0.51
LMNN_wHSV-DCHs	0.18	0.58	0.72	0.80
OMRR_wHSV	0.23	0.72	0.81	0.84
OMRR_DCHs	0.28	0.74	0.79	0.82
OMRR_wHSV-DCHs	**0.31**	**0.81**	**0.87**	**0.93**

Table 1. Performance comparison via Mean Reciprocal Rank (MRR). "-" means that the result is not available.

5. Conclusion and future work

In this paper, we argue that person re-identification models are better to directly optimize lisewise ranking criteria to meet practical demands, instead of treating it as a binary yes/no verification problem or a pairwise relevant/irrelevant classification problem. We present a generic maximum-margin based structured learning model for this purpose. Equipped with an efficient learning algorithm, we are able to optimizes the listwise ranking loss – MRR, which coincides with the widely used CMC performance evaluation criterion. Experiments on commonly used benchmark datasets show that the proposed model performs significantly better than other state-of-the-art approaches. Future work can be done on exploring new feature (besides color) and handling multi-shot re-identification cases.

Acknowledgments This work was supported by "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society", Special Coordination Fund for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

References

[1] S. Chakrabarti and R. Khanna. Structured learning for non-smooth ranking losses. In *In SIGKDD Conference*, pages 88–96. ACM, 2008.

[2] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In *ACCV (4)*, pages 501–512, 2010.

[3] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher. Appearance-based person reidentification in camera networks: problem overview and current approaches. *Journal of Ambient Intelligence and Humanized Computing*, 2:127–151, 2011.

[4] A. Ess, B. Leibe, and L. van Gool. Depth and appearance for mobile scene analysis. In *ICCV*, pages 1–8, 2007.

[5] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[6] D. Gray and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, pages 41–47, 2007.

[7] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, pages 262–275, 2008.

[8] K. E. Iverson. *A Programming Language*. New York: Wiley, 1962.

[9] T. Joachims. A support vector method for multivariate performance measures. In *ICML*, pages 377–384, 2005.

[10] T.-Y. Liu. Learning to rank for information retrieval. *Found. Trends Inf. Retr.*, 3:225–331, March 2009.

[11] B. McFee and G. Lanckriet. Metric learning to rank. In *ICML*, pages 775–782, 2010.

[12] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *BMVC*, 2010.

[13] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *SIBGRAPI*, pages 322–329, 2009.

[14] E. Voorhees. Proceedings of the 8th text retrieval conference, 1999. TREC-8 Question Answering Track Report.

[15] Y. Yue, T. Finley, F. Radlinski, and T. Joachims. A support vector method for optimizing average precision. In *ACM SIGIR*, pages 271–278, 2007.