Re-ranking Person Re-identification with k-reciprocal Encoding

Zhun Zhong[†], Liang Zheng[§], Donglin Cao[†], Shaozi Li[†]
[†]Xiamen University [§]University of Technology Sydney
{zhunzhong007,liangzheng06}@gmail.com {another,szlig}@xmu.edu.cn

Abstract

When considering person re-identification (re-ID) as a retrieval process, re-ranking is a critical step to improve its accuracy. Yet in the re-ID community, limited effort has been devoted to re-ranking, especially those fully automatic, unsupervised solutions. In this paper, we propose a k-reciprocal encoding method to re-rank the re-ID results. Our hypothesis is that if a gallery image is similar to the probe in the k-reciprocal nearest neighbors, it is more likely to be a true match. Specifically, given an image, a k-reciprocal feature is calculated by encoding its k-reciprocal nearest neighbors into a single vector, which is used for re-ranking under the Jaccard distance. The final distance is computed as the combination of the original distance and the Jaccard distance. Our re-ranking method does not require any human interaction or any labeled data, so it is applicable to large-scale datasets. Experiments on the large-scale Market-1501, CUHK03, MARS, and PRW datasets confirm the effectiveness of our method. Our code will be released soon.

1. Introduction

Person re-identification (re-ID) [50, 3, 23, 31, 27, 29] is a challenging task in computer vision. In general, re-ID can be regarded as a retrieval problem. Given a probe person, we want to search in the gallery for images containing the same person in a cross-camera mode. After an initial ranking list is obtained, a good practice consists of adding a re-ranking step, with the expectation that the relevant images will receive higher ranks. In this paper, we thus focus on the re-ranking issue.

Re-ranking has been mostly studied in generic instance retrieval [5, 14, 34, 35]. The main advantage of many re-ranking methods is that it can be implemented without requiring additional training samples, and that it can be applied to any initial ranking result.

The effectiveness of re-ranking depends heavily on the quality of the initial ranking list. A number of previous works exploit the similarity relationships between top-ranked images (such as the k-nearest neighbors) in the initial ranking list [5, 14, 34, 35, 43, 44]. An underlying assumption is that if a returned image ranks within the k-nearest neighbors of the probe, it is likely to be a true match which can be used for the subsequent re-ranking. Nevertheless, situation may deviate from optimal cases: false matches may well be included in the k-nearest neighbors of the probe. For example, in Fig. 1, P1, P2, P3 and P4 are four true matches to the probe, but all of them are not included in the top-4 ranks. We observe some false matches (N1-N6) receive high ranks. As a result, directly using the top-k ranked images may introduce noise in the re-ranking systems and compromise the final result.

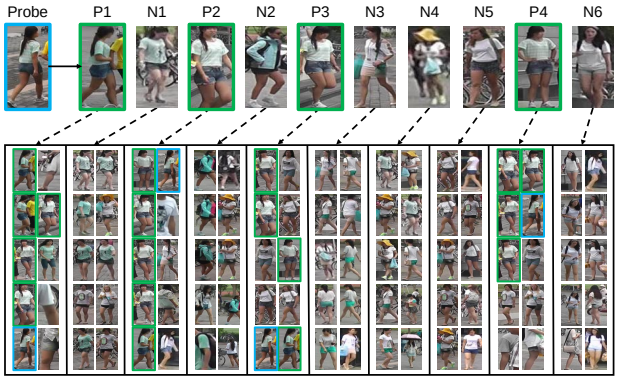

Figure 1. Illustration of the nearest neighborhoods of a person re-identification application. **Top:** The query and its 10-nearest neighbors, where P1-P4 are positives, N1-N6 are negatives. **Bottom:** Each two columns shows 10-nearest neighbors of the corresponding person. Blue and green box correspond to the probe and positives, respectively. We can observe that the probe person and positive persons are 10-nearest neighbors reciprocally.

In literature, the k-reciprocal nearest neighbor [14, 34] is an effective solution to the above-mentioned problem, *i.e.*, the pollution of false matches to the top-k images. When two images are called k-reciprocal nearest neighbors, they are both ranked top-k when the other image is taken as the probe. Therefore, the k-reciprocal nearest neighbor serves as a stricter rule whether two images are true matches or not. In Fig. 1, we observe that the probe is a reciprocal

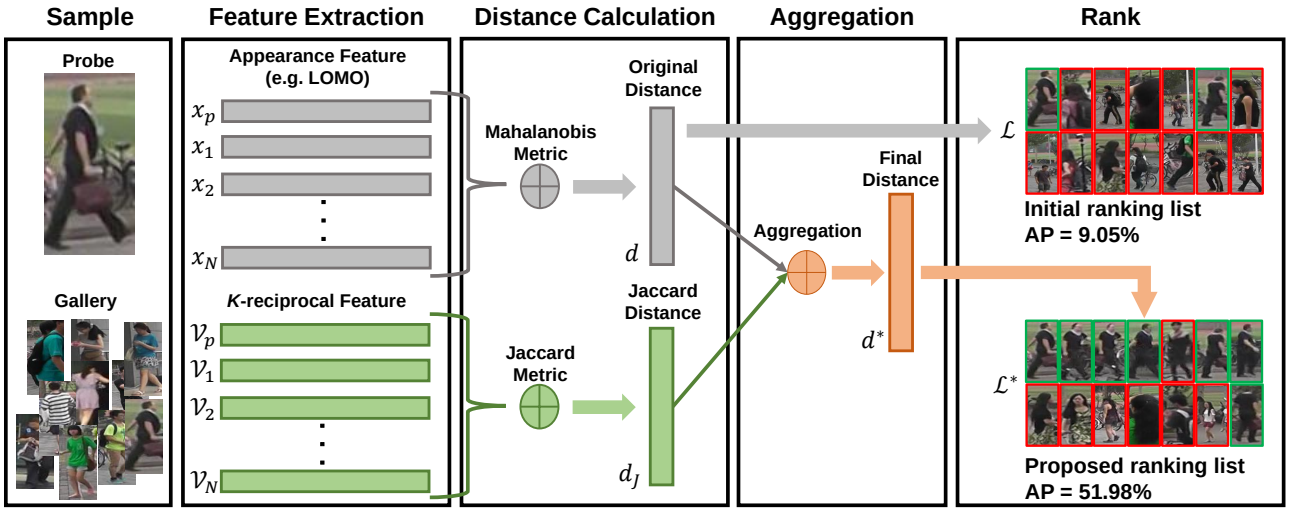

Figure 2. Proposed re-ranking framework for person re-identification. Given a probe p and a gallery, the appearance feature and k-reciprocal feature are extracted for each person. Then the original distance d and Jaccard distance d_J are calculated for each pair of the probe person and gallery person. The final distance d^* is computed as the combination of d and d_J, which is used to obtain the proposed ranking list.

neighbor to the true matched images, but not to the false matches. This observation identifies the true matches in the initial ranking list to improve the re-ranking results.

Given the above considerations, this paper introduces a k-reciprocal encoding method for re-ID re-ranking. Our approach consists of three steps. First, we encode the weighted k-reciprocal neighbor set into a vector to form the k-reciprocal feature. Then, the Jaccard distance between two images can be computed by their k-reciprocal features. Second, to obtain a more robust k-reciprocal feature, we develop a local query expansion approach to further improve the re-ID performance. Finally, the final distance is calculated as the weighted aggregation of the original distance and the Jaccard distance. It is subsequently used to acquire the re-ranking list. The framework of the proposed approach is illustrated in Fig. 2. To summarize, the contributions of this paper are:

- We propose a k-reciprocal feature by encoding the k-reciprocal feature into a singe vector. The re-ranking process can be easily performed by vector comparison.
- Our approach does not require any human interaction or annotated data, and can be applied to any person re-ID ranking result in an automatic and unsupervised way.
- The proposed method effectively improves the person re-ID performance on several datasets, including Market-1501, CUHK03, MARS, and PRW. In particular, we achieve the state-of-the-art accuracy on Market-1501 in both rank-1 and mAP.

2. Related Work

We refer the interested readers to [3, 50] for a detailed review of person re-identification (re-ID). Here we focus on research that aims at re-ranking methods for object retrieval and particularly for re-ID.

Re-ranking for object retrieval. Re-ranking methods have been successfully studied to improve object retrieval accuracy [51]. A number of works utilize the k-nearest neighbors to explore similarity relationships to address the re-ranking problem. Chum *et al.* [5] propose the average query expansion (AQE) method, where a new query vector is obtained by averaging the vectors in the top-k returned results, and is used to re-query the database. To take advantage of the negative sample which is far away from the query image, Arandjelović and Zisserman [1] develop the discriminative query expansion (DQE) to use a linear SVM to obtain a weight vector. The distance from the decision boundary is employed to revise the initial ranking list. Shen *et al.* [35] make use of the k-nearest neighbors of the initial ranking list as new queries to produce new ranking lists. The new score of each image is calculated depending on its positions in the produced ranking lists. More recently, sparse contextual activation (SCA) [2] propose to encode the neighbor set into a vector, and to indicate samples similarity by generalized Jaccard distance. To prevent the pollution of false matches to the top-k images, the concept of k-reciprocal nearest neighbors is adopted in [14, 34]. In [14], the contextual dissimilarity measure (CDM) is proposed to refine the similarity by iteratively regularizing the average distance of each point to its neighborhood. Qin *et al.* [34] formally present the concept of k-reciprocal nearest neighbors. The k-reciprocal nearest neighbors are con-

sidered as highly relevant candidates, and used to construct closed set for re-ranking the rest of dataset. Our work departs from both works in several aspects. We do not symmetrize nearest neighborhood relationship to refine the similarity as [14], or directly consider the k-reciprocal nearest neighbors as top ranked samples like [34]. Instead we calculate a new distance between two images by comparing their k-reciprocal nearest neighbors.

Re-ranking for re-ID. Most existing person re-identification methods mainly focus on feature representation [41, 12, 23, 48, 21] or metric learning [23, 17, 9, 32, 45]. Recently, several researchers [10, 33, 28, 24, 49, 20, 11, 19, 42, 44] have paid attention to re-ranking based method in the re-ID community. Li *et al.* [20] develop a re-ranking model by analyzing the relative information and direct information of near neighbors of each pair of images. In [11], an unsupervised re-ranking model is learnt by jointly considering the content and context information in the ranking list, which effectively remove ambiguous samples to improve the performance of re-ID. Leng *et al.* [19] propose a bidirectional ranking method to revise the initial ranking list with the new similarity computed as the fusion of both content and contextual similarity. Recently, the common nearest neighbors of different baseline methods are exploited to re-ranking task [42, 44]. Ye *et al.* [42] combine the common nearest neighbors of global and local features as new queries, and revise the initial ranking list by aggregating the new ranking lists of global and local features. In [44], the k-nearest neighbor set is utilized to calculate both similarity and dissimilarity from different baseline method, then the aggregation of similarity and dissimilarity is performed to optimize the initial ranking list. Continues progress of these mentioned methods in re-ranking promises to make future contributions to discovering further information from k-nearest neighbors. However, using the k-nearest neighbors to implement re-ranking directly may restrict the overall performance since false matches are often included. To tackle this problem, in this paper, we investigate the importance of k-reciprocal neighbors in person re-ID and hence design a simple but effective re-ranking method.

3. Proposed Approach

3.1. Problem Definition

Given a probe person p and the gallery set with N images $\mathcal{G} = \{g_i \mid i = 1, 2, ...N\}$, the original distance between two persons p and g_i can be measured by Mahalanobis distance,

$$d(p, g_i) = (x_p - x_{g_i})^\top \mathbf{M}(x_p - x_{g_i}) \tag{1}$$

where x_p and x_{g_i} represents the appearance feature of probe p and gallery g_i, respectively, and $\mathbf{M}$ is a positive semidefinite matrix.

The initial ranking list $\mathcal{L}(p, \mathcal{G}) = \{g_1^0, g_2^0, ...g_N^0\}$ can be obtained according to the pairwise original distance between probe p and gallery g_i, where $d(p, g_i^0) < d(p, g_{i+1}^0)$. Our goal is to re-rank $\mathcal{L}(p, \mathcal{G})$, so that more positive samples rank top in the list, and thus to improve the performance of person re-identification (re-ID).

3.2. K-reciprocal Nearest Neighbors

Following [34], we define $N(p, k)$ as the k-nearest neighbors (*i.e.* the top-k samples of the ranking list) of a probe p:

$$N(p, k) = \{g_1^0, g_2^0, ..., g_k^0\}, |N(p, k)| = k \tag{2}$$

where $|\cdot|$ denotes the number of candidates in the set. The k-reciprocal nearest neighbors $\mathcal{R}(p, k)$ can be defined as,

$$\mathcal{R}(p, k) = \{(g_i \in N(p, k)) \cap (p \in N(g_i, k))\} \tag{3}$$

According to the previous description, the k-reciprocal nearest neighbors are more related to probe p than k-nearest neighbors. However, due to variations in illuminations, poses, views and occlusions, the positive images may be excluded from the k-nearest neighbors, and subsequently not be included in the k-reciprocal nearest neighbors. To address this problem, we incrementally add the $\frac{1}{2}k$-reciprocal nearest neighbors of each candidate in $\mathcal{R}(p, k)$ into a more robust set $\mathcal{R}^*(p, k)$ according to the following condition

$$\begin{aligned} &\mathcal{R}^*(p, k) \leftarrow \mathcal{R}(p, k) \cup \mathcal{R}(q, \tfrac{1}{2}k) \\ &s.t. \left|\mathcal{R}(p, k) \cap \mathcal{R}(q, \tfrac{1}{2}k)\right| \geqslant \tfrac{2}{3}\left|\mathcal{R}(q, \tfrac{1}{2}k)\right|, \\ &\forall q \in \mathcal{R}(p, k) \end{aligned} \tag{4}$$

By this operation, we can add into $\mathcal{R}^*(p, k)$ more positive samples which are more similar to the candidates in $\mathcal{R}(p, k)$ than to the probe p. This is stricter against including too many negative samples compared to [34]. In Fig. 3, we show an example of the expansion process. Initially, the hard positive G is missed out in $\mathcal{R}(Q, 20)$. Interestingly, G is included in $\mathcal{R}(C, 10)$, which is beneficial information for bringing positive G back. Then, we can apply Eq. 4 to add G into $\mathcal{R}^*(Q, 20)$. Therefore, after expansion process, more positive samples could be added into $\mathcal{R}^*(p, k)$. Different from [34], we do not directly take the candidates in $\mathcal{R}^*(p, k)$ as top ranked images. Instead, we consider $\mathcal{R}^*(p, k)$ as contextual knowledge to re-calculate the distance between the probe and gallery.

3.3. Jaccard Distance

In this subsection, we re-calculate the pairwise distance between the probe p and the gallery g_i by comparing their k-reciprocal nearest neighbor set. As described earlier [2]

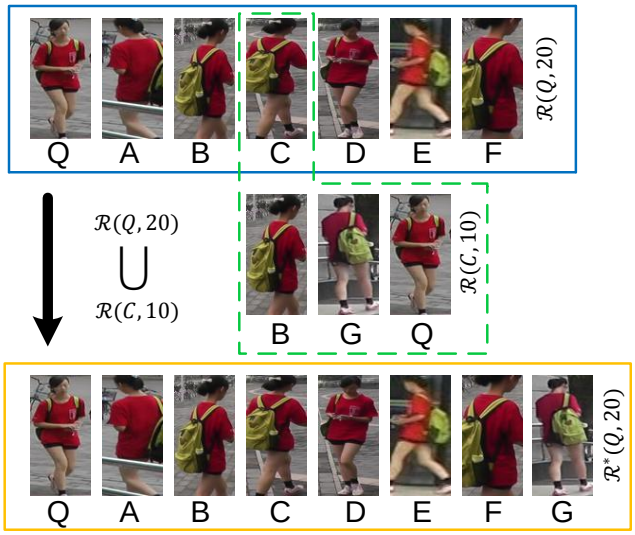

Figure 3. Example of the k-reciprocal neighbors expansion process. The positive person G which is similar to C is added into $\mathcal{R}^*(Q,20)$.

[44], we believe that if two images are similar, their k-reciprocal nearest neighbor sets overlap, *i.e.*, there are some duplicate samples in the sets. And the more duplicate samples, the more similar the two images are. The new distance between p and g_i can be calculated by the Jaccard metric of their k-reciprocal features as:

$$d_J(p,g_i) = 1 - \frac{|\mathcal{R}^*(p,k) \cap \mathcal{R}^*(g_i,k)|}{|\mathcal{R}^*(p,k) \cup \mathcal{R}^*(g_i,k)|} \tag{5}$$

where $|\cdot|$ denotes the number of candidates in the set. We adopt Jaccard distance to name this new distance. Although the above method could capture the similarity relationships between two images, there still remains three obvious shortcomings:

- It is very time-consuming to get the intersection and union of two neighbor sets $\mathcal{R}^*(p,k)$ and $\mathcal{R}^*(g_i,k)$ in many cases, and it becomes more challenging while the Jaccard distance is needed to be calculated for all image pairs. An alternative way is to encode the neighbor set into an easier but equivalent vector, reducing the computational complexity greatly, while maintaining original structure in neighbor set.
- The distance calculation method weighs all neighbors equally, leading to a simple but not discriminative neighbor set. In fact, neighbors that are closer to probe p are more likely to be true positives. Therefore, it is convincing and reasonable to re-calculate weights based on the original distance, and assign large weights to nearer samples.
- Simply taking the contextual information into account will pose considerable barriers when attempting to measure similarity between two persons, since unavoidable variation makes it difficult to discriminate sufficient contextual information. Hence, incorporating original distance and Jaccard distance becomes important for a robust distance.

To address the first two shortcomings, the k-reciprocal feature is proposed, by encoding the k-reciprocal nearest neighbor set into a vector $\mathcal{V}_p = [\mathcal{V}_{p,g_1}, \mathcal{V}_{p,g_2}, ..., \mathcal{V}_{p,g_N}]$, where $\mathcal{V}_{p,g_i}$ is initially defined by a binary indicator function as

$$\mathcal{V}_{p,g_i} = \begin{cases} 1 & \text{if } g_i \in \mathcal{R}^*(p,k) \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

In this way, the k-reciprocal neighbor set can be represented as an N-dimensional vector, with each item of the vector indicating whether the corresponding image is included in $\mathcal{R}^*(p,k)$. However, this function still consider each neighbor as equal. Intuitively, the neighbor who is closer to the probe p should be more similar with the probe p. Thus, we reassign weights according to the original distance between the probe and its neighbor, we redefine Eq. 6 by the Gaussian kernel of the pairwise distance as

$$\mathcal{V}_{p,g_i} = \begin{cases} \mathrm{e}^{-d(p,g_i)} & \text{if } g_i \in \mathcal{R}^*(p,k) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

In this way, the hard weighting (0 or 1) is converted into soft weighting, with closer neighbors assigned larger weights while farther neighbors smaller weights. Based on the above definition, the number of candidates in the intersection and union set can be calculated as

$$|\mathcal{R}^*(p,k) \cap \mathcal{R}^*(g_i,k)| = \|\min(\mathcal{V}_p, \mathcal{V}_{g_i})\|_1 \tag{8}$$

$$|\mathcal{R}^*(p,k) \cup \mathcal{R}^*(g_i,k)| = \|\max(\mathcal{V}_p, \mathcal{V}_{g_i})\|_1 \tag{9}$$

where $\min$ and $\max$ operate the element-based minimization and maximization for two input vectors. $\|\cdot\|_1$ is L_1 norm. Thus we can rewrite the Jaccard distance in Eq. 5 as

$$d_J(p,g_i) = 1 - \frac{\sum_{j=1}^{N} \min(\mathcal{V}_{p,g_j}, \mathcal{V}_{g_i,g_j})}{\sum_{j=1}^{N} \max(\mathcal{V}_{p,g_j}, \mathcal{V}_{g_i,g_j})} \tag{10}$$

By formula transformation from Eq. 5 to Eq. 10, we have succeed in converting the set comparison problem into pure vector calculation, which is much easier practically.

3.4. Local Query Expansion

Emulating the idea that the images from the same class may share similar features, we use the k-nearest neighbors of the probe p to implement the local query expansion. The local query expansion is defined as

$$\mathcal{V}_p = \frac{1}{|N(p,k)|} \sum_{g_i \in N(p,k)} \mathcal{V}_{g_i} \tag{11}$$

As a result, the k-reciprocal feature $\mathcal{V}_p$ is expanded by the k-nearest neighbors of probe p. Note that, we implement this query expansion both on the probe p and galleries g_i. Since there will be noise in the k-nearest neighbors, we limit the size of $N(p,k)$ used in the local query expansion to a smaller value. In order to distinguish between the size of $\mathcal{R}^*(g_i,k)$ and $N(p,k)$ used in Eq. 7 and Eq. 11, we denote the former as k_1 and the latter as k_2, respectively, where $k_1 > k_2$.

3.5. Final Distance

In this subsection, we focus on the third shortcoming of Eq. 5. While most existing re-ranking methods ignore the importance of original distance in re-ranking, we jointly aggregate the original distance and Jaccard distance to revise the initial ranking list, the final distance d^* is defined as

$$d^*(p,g_i) = (1-\lambda)d_J(p,g_i) + \lambda d(p,g_i) \tag{12}$$

where $\lambda \in [0,1]$ denotes the penalty factor, it penalizes galleries far away from the probe p. When $\lambda = 0$, only the k-reciprocal distance is considered. On the contrary, when λ = 1, only the original distance is considered. The effect of λ is discussed in section 4. Finally, the revised ranking list $\mathcal{L}^*(p,\mathcal{G})$ can be obtained by ascending sort of the final distance.

3.6. Complexity Analysis

In the proposed method, most of the computation costs focus on pairwise distance computing for all gallery pairs. Suppose the size of the gallery set is N, the computation complexity required for the distance measure and the ranking process is $O(N^2)$ and $O(N^2logN)$, respectively. However, in practical applications, we can calculate the pairwise distance and obtain the ranking lists for the gallery in advance offline. As a result, given a new probe p, we only need to compute the pairwise distance between p and gallery with computation complexity $O(N)$ and to rank all final distance with computation complexity $O(NlogN)$.

4. Experiments

4.1. Datasets and Settings

Datasets Because our re-ranking approach is based on the comparison of similar neighbors between two persons, we conducted experiments on four large-scale person re-identification (re-ID) benchmark datasets that contain multiple positive samples for each probe in the gallery : including two image-based datasets, Market-1501 [48], CUHK03 [22] , a video-based dataset MARS [47], and an end-to-end dataset PRW [52] (see Table 1 for an overview).

Market-1501 [48] is currently the largest image-based re-ID benchmark dataset. It contains 32,668 labeled bounding boxes of 1,501 identities captured from 6 different view points. The bounding boxes are detected using Deformable Part Model (DPM) [8]. The dataset is split into two parts: 12,936 images with 751 identities for training and 19,732 images with 750 identities for testing. In testing, 3,368 hand-drawn images with 750 identities are used as probe set to identify the correct identities on the testing set. We report the single-query evaluation results for this dataset.

Table 1. The details of datasets used in our experiments.

Datasets	# ID	# box	# box/ID	# cam
Market-1501 [48]	1,501	32,643	19.9	6
CUHK03 [22]	1,360	13,164	9.7	2
MARS [47]	1,261	1,067,516	13.2	6
PRW [52]	932	34,304	36.8	6

CUHK03 [22] is another large scale image-based dataset, which contains 13,164 images of 1,360 identities. Each identity is captured from two cameras in the CUHK campus, and has an average of 4.8 images in each camera. The dataset provides both manually labeled bounding boxes and DPM-detected bounding boxes. Since a real-world re-ID system has to rely on a person detector, the latter version of the data is ideal for testing performance given detector errors. In this paper, both experimental results on 'labeled' and 'detected' data are presented. Following the protocol in [22], we split the dataset into a training set consist of 1,160 identities and a testing set consist of 100 identities, and repeat 20 times for evaluation. The average result over all tests is reported.

MARS [47] is the largest video-based re-ID benchmark dataset to date, containing 1,261 identities and around 20,000 video sequences. These sequences are collected from 6 different cameras and each identity has 13.2 sequences on average. Each sequence is automatically obtained by the DPM as pedestrian detector and the GMMCP [6] as tracker. In addition, the dataset also contains 3,248 distractor sequences. The dataset is fixedly split into training and test sets, with 631 and 630 identities, respectively. In testing, 2,009 probes are selected for query.

PRW [52] is an end-to-end large-scale dataset. It is composed of 11,816 frames of 932 identities captured from six different cameras. A total of 43,110 annotated person bounding boxes are generated from these frames. Given a query bounding box, the dataset aims to first perform pedestrian detection on the raw frames to generate the gallery, and identify the correct bounding boxes from the gallery. The dataset is divided into a training set with 5,704 frames of 482 identities and a test set with 6,112 frames of 450 identities. In testing, 2,057 query images for 450 identities are selected for evaluation. A detected bounding box is considered correct if its IoU value with the ground truth is above 0.5.

Evaluation metrics We use two evaluation metrics to evaluate the performance of re-ID methods on all datasets.

The first one is the Cumulated Matching Characteristics (CMC). Considering re-ID as a ranking problem, we report the cumulated matching accuracy at rank-1. The other one is the mean average precision (mAP) considering re-ID as an object retrieval problem, as described in [48].

Feature representations The Local Maximal Occurrence (LOMO) features are used to represent the person appearance [23]. The LOMO extractor generate a 26,960-dimensional feature for each image. This feature is robust to view changes and illumination variations by concatenating the maximal pattern of joint HSV histogram and SILTP descriptor, and is also discriminative, by capturing local region characteristics of a person. In addition, the ID-discriminative Embedding (IDE) feature proposed in [52] is used. The IDE extractor is effectively trained on classification model including CaffeNet [18] and ResNet-50 [13]. It generates a 1,024-dim (or 2,048-dim) vector for each image, which is effective in large-scale re-ID datasets. For the convenience of description, we abbreviate the IDE trained on CaffeNet and ResNet-50 to IDE (C) and IDE (R) respectively. We use these two methods as the baseline of our re-id framework.

4.2. Experiments on Market-1501

We first evaluate our method on the largest image-based re-ID dataset. In this dataset, in addition to using LOMO and IDE features, we also use the BOW [48] feature. We trained the IDE feature on CaffeNet [18] and ResNet-50 [13]. We set k_1 to 20, k_2 to 6, and λ to 0.3. Results among various methods with our method are shown in Table 2. Our method consistently improves the rank-1 accuracy and mAP with all features, even with the IDE (R) which is trained on the powerful ResNet-50 model. Our method gains 3.06% improvement in rank-1 accuracy and significant 13.99% improvement in mAP for IDE (R). Comparing with two popular re-ranking methods, average query expansion (AQE) [5] and contextual dissimilarity measure (CDM) [14], our method outperforms them both in rank-1 accuracy and mAP. Moreover, experiments conducted with two metrics, KISSME [17] and XQDA [23] verify the effectiveness of our method on different distance metrics.

Table 3 compares the performance of our best approach, IDE (R) + KISSME + ours, with other state-of-the-art methods. Our best method impressively outperforms the previous work and achieves large margin advances compared with the state-of-the-art results in rank-1 accuracy, particularly in mAP.

4.3. Experiments on CUHK03

We evaluate our experiments on two settings, single-shot, and multi-shot. Following the protocol in [22], the single-shot setting randomly selects one image for each identity from each camera. In multi-shot setting, for each identity, we randomly select the query image from one camera, and use all images in another camera to construct the gallery set. We set k_1 to 7, k_2 to 3, and λ to 0.85 for single-shot, and k_1 to 10, k_2 to 5, and λ to 0.85 for multi-shot, respectively. Results for single-shot are shown in Table 4. As we can see that, when using IDE feature, our re-ranking results are almost equivalent to raw results. It is reasonable that our approach does not work. Since there is only one positive for each identity in the gallery, our approach could

Table 2. Comparison among various methods with our re-ranking approach on the Market-1501 dataset.

Method	Rank 1	mAP
BOW	35.84	14.75
BOW + Ours	39.85	19.90
BOW + KISSME	42.90	19.41
BOW + KISSME + Ours	44.77	25.64
BOW + XQDA	41.39	19.72
BOW + XQDA + Ours	42.61	24.98
LOMO + KISSME	41.12	19.02
LOMO + KISSME + Ours	45.22	28.44
LOMO + XQDA	43.56	22.44
LOMO + XQDA + Ours	48.34	32.21
IDE (C)	55.87	31.34
IDE (C) + AQE [5]	57.69	35.25
IDE (C) + CDM [14]	58.02	34.54
IDE (C) + Ours	58.79	42.06
IDE (C) + XQDA	57.72	35.95
IDE (C) + XQDA + Ours	61.25	46.79
IDE (C) + KISSME	58.61	35.40
IDE (C) + KISSME + Ours	61.82	46.81
IDE (R)	72.54	46.00
IDE (R) + AQE [5]	73.20	50.14
IDE (R) + CDM [14]	73.66	49.53
IDE (R) + Ours	74.85	59.87
IDE (R) + XQDA	71.41	48.89
IDE (R) + XQDA + Ours	75.14	61.87
IDE (R) + KISSME	73.60	49.05
IDE (R) + KISSME + Ours	**77.11**	**63.63**

Table 3. Comparison of our method with state-of-the-art on the Market-1501 dataset.

Method	Rank 1	mAP
SDALF [7]	20.53	8.20
eSDC [46]	33.54	13.54
BOW [48]	34.40	14.09
PersonNet [40]	37.21	18.57
dCNN [36]	39.40	19.60
LOMO + XQDA [23]	43.79	22.22
MSTCNN [26]	45.10	-
WARCA [15]	45.16	-
MBCNN [37]	45.58	26.11
HistLBP+kLFDA [16]	46.50	-
TMA [30]	47.92	22.31
DLDA [39]	48.15	29.94
CAN [25]	48.24	24.43
SCSP [4]	51.90	26.35
DNS [45]	61.02	35.68
Gated [38]	65.88	39.55
IDE (R) + KISSME + Ours	**77.11**	**63.63**

Table 4. Comparison among various methods with our re-ranking approach on the CUHK03 dataset in single-shot setting.

Method	Labeled		Detected	
	Rank 1	mAP	Rank 1	mAP
LOMO + XQDA [23]	49.7	56.4	44.6	51.5
LOMO + XQDA + Ours	50.0	56.8	45.9	52.6
IDE (C) [52]	57.0	63.1	54.1	60.4
IDE (C) + Ours	57.2	63.2	54.2	60.5
IDE (C) + XQDA [52]	61.7	67.6	58.9	64.9
IDE (C) + XQDA + Ours	61.6	67.6	58.5	64.7

Table 5. Comparison among various methods with our re-ranking approach on the CUHK03 dataset in multi-shot setting.

Method	Labeled		Detected	
	Rank 1	mAP	Rank 1	mAP
LOMO + XQ. [23]	57.4	50.7	52.0	45.1
LOMO + XQDA + Ours	59.8	58.0	52.9	51.9
IDE (C) [52]	64.7	59.0	60.8	55.5
IDE (C) + Ours	64.9	63.2	61.5	59.7
IDE (C) + XQDA [52]	68.6	63.8	64.4	60.3
IDE (C) + XQDA + Ours	69.1	68.4	64.6	64.5
IDE (R) + XQDA [52]	69.80	66.16	69.62	65.75
IDE (R) + XQDA + Ours	**69.90**	**70.89**	**69.67**	**72.45**

not obtain sufficient contextual information. Even so, our approach gains nearly 1% improvement for rank-1 accuracy and mAP while applying LOMO feature on both 'labeled' and 'detected' setting, except LOMO + XQDA in 'labeled' setting. Experiments show that, in the case of single-shot, our method does no harm to results, and has the chance to improve the performance. To further demonstrate the effectiveness of our method, we conduct experiments in multi-shot setting. Results in Table 5 show that, in all cases, our method significantly improves mAP, and also slightly improves rank-1 accuracy. Especially for LOMO + XQDA, our method gains an increase of 2.4% in rank-1 accuracy and 8.7% in mAP on 'labeled' setting.

4.4. Experiments on MARS

We also evaluate our method on video-based dataset. On this dataset, we employ two features as the baseline methods, LOMO and IDE. For each sequence, we first extract feature for each image, and use max pooling to combine all features into a fixed-length vector. We set k_1 to 20, k_2 to 6, and λ to 0.3 in this dataset. The performance of our method on different features and metrics are reported in Table 6. As we can see, our re-ranking method consistently improves the rank-1 accuracy and mAP of the two different features. Results compared with average query expansion (AQE) [5] and contextual dissimilarity measure (CDM) [14] show our method outperforms them in both rank-1 accuracy and mAP. Moreover, our method can even improve the rank-1 accuracy and mAP in all cases while discriminative metrics are used. In particular, our method (IDE (R) + XQDA + Ours), based on the IDE (R) + XQDA proposed in [47], improves the rank-1 accuracy from 70.51% to 73.94% and the mAP from 55.12% to 68.45%. Experimental results demonstrate that our re-ranking method is also effective on video-based re-ID problem. We believe that results of this problem will be further improved by combining more sophisticated feature model with our method.

Table 6. Comparison among various methods with our re-ranking approach on the MARS dataset.

Method	Rank 1	mAP
LOMO + KISSME [17]	30.86	15.36
LOMO + KISSME + Ours	31.31	22.38
LOMO + XQDA [23]	31.82	17.00
LOMO + XQDA + Ours	33.99	23.20
IDE (C) [47]	61.72	41.17
IDE (C) + AQE [5]	61.83	47.02
IDE (C) + CDM [14]	62.05	44.23
IDE (C) + Ours	62.78	51.47
IDE (C) + KISSME [47]	65.25	44.83
IDE (C) + KISSME + Ours	66.87	56.18
IDE (C) + XQDA [47]	65.05	46.87
IDE (C) + XQDA + Ours	67.78	57.98
IDE (R)	62.73	44.07
IDE (R) + AQE [5]	63.74	49.14
IDE (R) + CDM [14]	64.11	47.68
IDE (R) + Ours	65.61	57.94
IDE (R) + KISSME	70.35	53.27
IDE (R) + KISSME + Ours	72.32	67.29
IDE (R) + XQDA	70.51	55.12
IDE (R) + XQDA + Ours	**73.94**	**68.45**

Table 7. Comparison among various methods with our re-ranking approach on the PRW dataset.

Method	Rank 1	mAP
LOMO + XQDA [23]	34.9	13.4
LOMO + XQDA + Ours	37.1	19.2
IDE (C)	51.03	25.09
IDE (C) + Ours	**52.54**	**31.51**

4.5. Experiments on PRW

We also evaluate our method on the end-to-end re-ID dataset. This dataset is more challenging than image-based and video-based datasets, since it requires to detect person from a raw image and identify the correct person from the detected galleries. Following [52], we first use DPM to detect candidate bounding boxes of persons on a large raw image, and then query on the detected bounding boxes. We use LOMO and IDE to extract features for each bounding box, and take these two methods as baselines. We set k_1 to 20, k_2 to 6, and λ to 0.3. Experiment results are shown in Table 7. It can be seen that, our method consistently improves the rank-1 accuracy and mAP of both LOMO and IDE feature, demonstrating that our method is effective on end-to-end re-ID task.

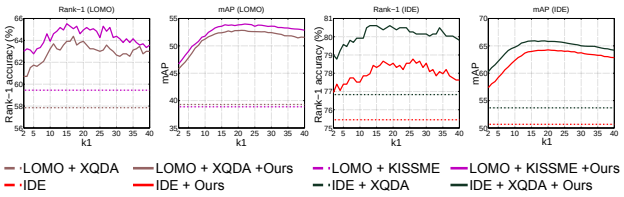

Figure 4. The impact of the parameter k_1 on re-ID performance on the Market-1501 dataset. We fix the k_2 at 6 and λ at 0.3.

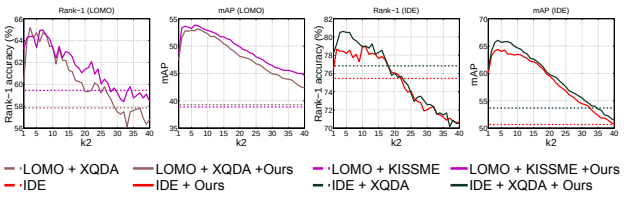

Figure 5. The impact of the parameter k_2 on re-ID performance on the Market-1501 dataset. We fix the k_1 at 20 and λ at 0.3.

4.6. Parameters Analysis

The parameters of our method are analyzed in this subsection. The baseline methods are LOMO [23] and IDE [52] trained on CaffeNet. We evaluate the influence of k_1, k_2, and λ on rank-1 accuracy and mAP on the Market-1501 dataset. To conduct experimental analyses, we randomly split the original training set into training and validation sets, with 425 and 200 identities respectively.

Fig. 4 shows the impact of the size of k-reciprocal neighbors set on rank-1 accuracy and mAP. It can be seen that, our method consistently outperforms the baselines both on the rank-1 accuracy and mAP with various values of k_1. The mAP first increases with the growth of k_1, and then begins a slow decline after k_1 surpasses a threshold. Similarly, as k_1 grows, the rank-1 accuracy first rises with fluctuations; and after arriving at the optimal point around $k_1 = 20$, it starts to drop. With a too large value of k_1, there will be more false matches included in the k-reciprocal set, resulting in a decline in performance.

The impact of k_2 are shown in Fig. 5. When k_2 is equal to 1, the local query expansion is not considered. Obviously, the performance grows as k_2 increases in a reasonable range. Notice that, assigning a much too large value to k_2 reduces the performance. Since it may lead to exponentially containing false matches in local query expansion, which undoubtedly harm the feature and thus the performance. As a matter of fact, the local query expansion is very beneficial for further enhancing the performance when setting an appropriate value to k_2.

The impact of the parameter λ is shown in Fig. 6. Notice that, when λ is set to 0, we only consider the Jaccard distance as the final distance; in contrast, when λ equal to 1, the Jaccard distance is left out, and the result is exactly the baseline result obtained using pure original distance. It can be observed that when only Jaccard distance is considered, our method consistently outperforms the baseline. This demonstrates that the proposed Jaccard distance is effective for re-ranking. Moreover, when simultaneously considering both the original distance and the Jaccard distance, the performance obtains a further improvement when the value of λ is around 0.3, demonstrating that the original distance is also important for re-ranking.

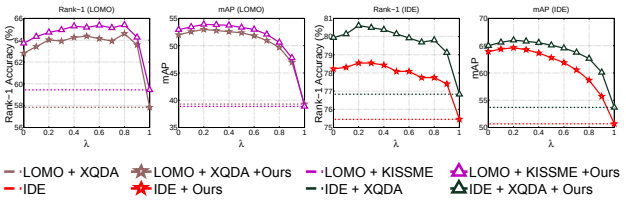

Figure 6. The impact of the parameter λ on re-ID performance on the Market-1501 dataset. We fix the k_1 at 20 and k_2 at 6.

Figure 7. Example results of four probes on the Market-1501 dataset. For each probe, the first row and the second correspond to the ranking results produced by IDE and IDE + Ours, respectively. Person surrounded by green box denotes the same person as the probe.

In Fig. 7, four example results are shown. The proposed method, IDE + Ours, effectively ranks more true persons in the top of ranking list which are missed in the ranking list of IDE.

5. Conclusion

In this paper, we address the re-ranking problem in person re-identification (re-ID). We propose a k-reciprocal feature by encoding the k-reciprocal nearest neighbors into a

single vector, thus the re-ranking process can be readily performed by vector comparison. To capture the similarity relationships from similar samples, the local expansion query is proposed to obtain a more robust k-reciprocal feature. The final distance based on the combination of the original distance and Jaccard distance produces effective improvement of the re-ID performance on several large-scale datasets. It is worth mentioning that our approach is fully automatic and unsupervised, and can be easily implemented to any ranking result.

References

[1] R. Arandjelović and A. Zisserman. Three things everyone should know to improve object retrieval. In *CVPR*, 2012. 2

[2] S. Bai and X. Bai. Sparse contextual activation for efficient visual re-ranking. *IEEE TIP*, 2016. 2, 3

[3] A. Bedagkar-Gala and S. K. Shah. A survey of approaches and trends in person re-identification. *Image and Vision Computing*, 2014. 1, 2

[4] D. Chen, Z. Yuan, B. Chen, and N. Zheng. Similarity learning with spatial constraints for person re-identification. In *CVPR*, 2016. 6

[5] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *ICCV*, 2007. 1, 2, 6, 7

[6] A. Dehghan, S. Modiri Assari, and M. Shah. Gmmcp tracker: Globally optimal generalized maximum multi clique problem for multiple object tracking. In *CVPR*, 2015. 5

[7] 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

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE TPAMI*, 2010. 5

[9] J. García, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, and C. Micheloni. Modeling feature distances by orientation driven classifiers for person re-identification. *Elsevier VCIP*, 2016. 3

[10] J. Garcia, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, and C. Micheloni. Discriminant context information analysis for post-ranking person re-identification. *IEEE TIP*, 2017. 3

[11] J. Garcia, N. Martinel, C. Micheloni, and A. Gardel. Person re-identification ranking optimisation by discriminant context information analysis. In *ICCV*, 2015. 3

[12] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, 2008. 3

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 6

[14] H. Jegou, H. Harzallah, and C. Schmid. A contextual dissimilarity measure for accurate and efficient image search. In *CVPR*, 2007. 1, 2, 3, 6, 7

[15] C. Jose and F. Fleuret. Scalable metric learning via weighted approximate rank component analysis. In *ECCV*, 2016. 6

[16] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke. A comprehensive evaluation and benchmark for person re-identification: Features, metrics, and datasets. *arXiv preprint arXiv:1605.09653*, 2016. 6

[17] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012. 3, 6, 7

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012. 6

[19] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen. Person re-identification with content and context re-ranking. *Springer MTAP*, 2015. 3

[20] W. Li, Y. Wu, M. Mukunoki, and M. Minoh. Common-near-neighbor analysis for person re-identification. In *ICIP*, 2012. 3

[21] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 3

[22] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 5, 6

[23] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, 2015. 1, 3, 6, 7, 8

[24] C. Liu, C. Change Loy, S. Gong, and G. Wang. Pop: Person re-identification post-rank optimisation. In *ICCV*, 2013. 3

[25] H. Liu, J. Feng, M. Qi, J. Jiang, and S. Yan. End-to-end comparative attention networks for person re-identification. *arXiv preprint arXiv:1606.04404*, 2016. 6

[26] J. Liu, Z.-J. Zha, Q. Tian, D. Liu, T. Yao, Q. Ling, and T. Mei. Multi-scale triplet cnn for person re-identification. In *ACM MM*, 2016. 6

[27] A. J. Ma, J. Li, P. C. Yuen, and P. Li. Cross-domain person reidentification using domain adaptation ranking svms. *IEEE TIP*, 2015. 1

[28] A. J. Ma and P. Li. Query based adaptive re-ranking for person re-identification. In *ACCV*, 2014. 3

[29] A. J. Ma, P. C. Yuen, and J. Li. Domain transfer support vector ranking for person re-identification without target camera label information. In *ICCV*, 2013. 1

[30] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury. Temporal model adaptation for person re-identification. In *ECCV*, 2016. 6

[31] N. Martinel, G. L. Foresti, and C. Micheloni. Person reidentification in a distributed camera network framework. *IEEE transactions on cybernetics*, 2016. 1

[32] N. Martinel, C. Micheloni, and G. L. Foresti. Kernelized saliency-based person re-identification through multiple metric learning. *IEEE TIP*, 2015. 3

[33] V.-H. Nguyen, T. D. Ngo, K. M. Nguyen, D. A. Duong, K. Nguyen, and D.-D. Le. Re-ranking for person re-identification. In *SoCPaR*. IEEE, 2013. 3

[34] D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. Van Gool. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR*, 2011. 1, 2, 3

[35] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In *CVPR*, 2012. 1, 2

[36] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian. Deep attributes driven multi-camera person re-identification. In *ECCV*, 2016. 6

[37] E. Ustinova, Y. Ganin, and V. Lempitsky. Multi-region bilinear convolutional neural networks for person re-identification. *arXiv preprint arXiv:1512.05300*, 2015. 6

[38] R. R. Varior, M. Haloi, and G. Wang. Gated siamese convolutional neural network architecture for human re-identification. In *ECCV*, 2016. 6

[39] L. Wu, C. Shen, and A. v. d. Hengel. Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. *arXiv preprint arXiv:1606.01595*, 2016. 6

[40] L. Wu, C. Shen, and A. v. d. Hengel. Personnet: Person re-identification with deep convolutional neural networks. *arXiv preprint arXiv:1601.07255*, 2016. 6

[41] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *ECCV*, 2014. 3

[42] M. Ye, J. Chen, Q. Leng, C. Liang, Z. Wang, and K. Sun. Coupled-view based ranking optimization for person re-identification. In *MMM*. Springer, 2015. 3

[43] M. Ye, C. Liang, Z. Wang, Q. Leng, and J. Chen. Ranking optimization for person re-identification via similarity and dissimilarity. In *ACM MM*, 2015. 1

[44] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, and R. Hu. Person re-identification via ranking aggregation of similarity pulling and dissimilarity pushing. *IEEE TMM*, 2016. 1, 3

[45] L. Zhang, T. Xiang, and S. Gong. Learning a discriminative null space for person re-identification. In *CVPR*, 2016. 3, 6

[46] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013. 6

[47] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian. Mars: A video benchmark for large-scale person re-identification. In *ECCV*, 2016. 5, 7

[48] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 3, 5, 6

[49] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian. Query-adaptive late fusion for image search and person re-identification. In *CVPR*, 2015. 3

[50] L. Zheng, Y. Yang, and A. G. Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016. 1, 2

[51] L. Zheng, Y. Yang, and Q. Tian. Sift meets cnn: a decade survey of instance retrieval. *arXiv preprint arXiv:1608.01807*, 2016. 2

[52] L. Zheng, H. Zhang, S. Sun, M. Chandraker, and Q. Tian. Person re-identification in the wild. *arXiv preprint arXiv:1604.02531*, 2016. 5, 6, 7, 8