Ranking Optimization for Person Re-identification via Similarity and Dissimilarity

Mang Ye¹, Chao Liang¹,², Zheng Wang¹, Qingming Leng¹, Jun Chen¹,²
¹National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, 430072, China {yemang,cliang,wangzwhu,chenj}@whu.edu.cn
²Collaborative Innovation Center of Geospatial Technology, Wuhan, China. lenglengqingming@126.com

ABSTRACT
Person re-identification is a key technique to match different persons observed in non-overlapping camera views. Many researchers treat it as a special object retrieval problem, where ranking optimization plays an important role. Existing ranking optimization methods utilize the similarity relationship between the probe and gallery images to optimize the original ranking list in which dissimilarity relationship is seldomly investigated. In this paper, we propose to use both similarity and dissimilarity cues in a ranking optimization framework for person re-identification. Its core idea is based on the phenomenon that the true match should not only be similar to the strong similar samples of the probe but also dissimilar to the strong dissimilar samples. Extensive experiments have shown the great superiority of the proposed ranking optimization method.

Categories and Subject Descriptors
J.m [Computer Applications]: Miscellaneous;
H.3.3 [Information Systems]: INFORMATION STORAGE AND RETRIEVAL—Retrieval models

Keywords
Person Re-Identification, ranking optimization, similarity and dissimilarity

1. INTRODUCTION
In recent years, person re-identification (re-id) problem, namely matching people across disjoint camera views in a multi-camera system, has aroused an increasing interest in the multimedia analysis community [1]. The main challenges in person re-identification can be attributed to the significant visual changes in pose, illumination and viewpoint, making intra-personal variations even larger than that of inter-personal variations. Moreover, background clutters and occlusions cause additional difficulties.

To tackle the above problems, existing person re-id methods can be roughly divided into three categories [16]: appearance representation methods, metric learning methods, and ranking optimization methods. Appearance representation methods [2,4,11,19] aim to seek distinctive descriptions by utilizing the human body structure and position. Metric learning methods [5,8,12,21] attempt to learn a proper distance metric to obtain an accurate similarity measurement to reduce the differences in disjoint camera images. Ranking optimization methods [6,7,9,13,17] try to refine the original ranking lists by establishing the similarity relationships lying in the identification results. While the ranking optimization methods based on ranking orders do not rely on the similarity scores, which own more flexibility and popularity, thus can be easily reproduced [10]. Therefore, we focus on order-based ranking optimization method in this paper.

Related ranking optimization methods can be categorized as automotive re-ranking and interactive relevance feedback. Li et al. [7] analyzed the commonness of the near neighbors...
to optimize the issue. Leng et al. [6] proposed a automatic bidirectional ranking method based on content and context similarities. In this paper, we utilize the similarity relationships among samples to optimize the original ranking list with a common phenomenon. Both the similarity and dissimilarity cues are considered for person re-identification. A latent assumption is that true match should not only be similar to the strongly similar samples but also dissimilar to the strongly dissimilar samples of the probe person.

Motivated by the basic principles in social networks [6], good friends always tend to have more common friends. In other words, if a person is strongly similar to the friends (strongly similar samples) of the probe person, he will be more likely to be a friend of the probe, we name this phenomenon as conductive similarity. On the other hand, if a person is very similar to the strangers (strongly dissimilar samples) of the probe person, it is more prone to be a dissimilar sample of the probe person, we name it as insulative dissimilarity. An illustration example is shown in Fig. 1. Furthermore, an objective statistic data is conducted by a preliminary experiment as shown in Fig.2, the top-10 results are treated as strongly similar samples while the bottom-10 as strongly dissimilar samples. As a reference, the neutral samples are denoted by the middle-10 samples. For each probe, we compute the pair-wise distance between the true match (groundtruth) and other gallery images. The average distance of the true match to the strongly similar samples is reported as "Avg_sim", while "Avg_neu" and "Avg_dis" represents the distance to neutral samples and strongly dissimilar samples, respectively. As can be seen from the figure, the true match are much likely to be similar to the strongly similar samples and dissimilar to strongly dissimilar samples. Therefore, it is reasonable to optimize the original ranking list based on the conductive similarity and insulative dissimilarity.

In this paper, a ranking optimization method for person re-identification is presented. More specifically, the quasi-similar samples (which are similar to strongly similar samples) are pulled while the quasi-dissimilar samples (which are similar to strongly dissimilar samples) are pushed to optimize the original ranking list. The main contribution of this paper can be summarized as follows: (1) We present a combination of similarity and dissimilarity cues while dissimilarity is seldom conducted in person re-identification. (2) The ranking optimization is conducted based on ranking orders which can be easily applied to other method. We have validated our approach on two public datasets, the VIPeR [3] and CUHK01 [14]. The experimental results illustrate the effectiveness of our method and the optimized results compare favorably with state-of-the-art methods.

2. OUR APPROACH

2.1 Similarity Ranking Optimization

Similarity ranking optimization is divided into three steps: First, the strongly similar samples are obtained by the top-k results achieved by an original method. Second, the strongly similar samples are treated as new probes to re-query in the original gallery set, named backward re-query [6]. For each strongly similar sample, the weighted Jaccard similarity coefficient of two neighborhood image sets is computed as the new similarity to the probe image. After that, those quasi-similar samples are pulled based on their frequencies and a refined ranking list is generated. The details are discussed as follows.

In order to facilitate the following discussion, we denote the probe image as $p$ and the gallery image set as $G = \{g_i\mid i = 1, 2, \ldots, n\}$, where $n$ is the number of images in gallery set. An initial ranking list is achieved for a query $p$, and the top-$k$ ($k_+$) matching results are defined as the strongly similar samples, $N_+(p) = \{g_j\mid j = 1, 2, \ldots k_+\}$. After achieving $k_+$ strongly similar samples, we treat each $g_j^+$ in $N_+(p)$ as a new probe to search in the original gallery set. A new ranking list is obtained for each $g_j^+$. Similar to [6], images of the same person are intended to have more common $k$-nearest neighbors than that of different persons. Therefore, a weighted Jaccard similarity between the neighborhoods of $p$ and $g_j^+$ is computed by:

$$\text{Sim}(p, g_j^+) = \frac{w_0 \cdot r(p, g_j^+) \cdot |N_{k_+}(p) \cap N_{k_+}(g_j^+)|}{|N_{k_+}(p) \cup N_{k_+}(g_j^+)|} \quad (1)$$

where $|\cdot|$ denotes the cardinality and $N_{k_+}(g_j^+)$ is the $k$-nearest neighbours of $g_j^+$ achieved by the backward requery. And first item is a smoothing coefficient related to the original rank in $N_+(p)$. The decay factor is define as $w_0$ and we set $w_0 = 0.8$ as [18] in all experiments. $r(p, g_j^+)$ represents the rank of $g_j^+$ in the original ranking list and expresses the original similarity.

Above procedure mainly focus on optimizing the ranking orders of the strongly similar samples, and then the whole ranking lists order can be revised via conductive similarity. The detail is shown as following, after the backward requery, we’ll get $k_+$ ranking lists, and each top-$k$ results of the backward ranking list is expressed by $N_{k_+}(g_j^+)$. The quasi-similar set $Q_+(p)$ is denoted by the union set of top-$k$ results $Q_+(p) = \{N_{k_+}(g_j^+) \mid j = 1, 2, \ldots, k_+\}$. The weighted penalization of each quasi-similar sample $q_+$ is denoted by $w'_s(q_+)$, which is defined by its frequency and rank in $Q_+(p)$, which is formulated by:

$$w'_s(q_+) = \sum_{k=1}^{k_+} \log \{r(p, g_j^+) \cdot r(q_+, g_j)\} \quad (2)$$

where $r(p, g_j^+)$ denotes the original rank of $g_j^+$ in the original ranking list and $r(q_+, g_j)$ represents the rank of $q_+$ in $N_{k_+}(g_j^+)$. Note that if $q_+$ is not an element of $N_{k_+}(g_j^+)$, we set $r(q_+, g_j) = k_+$. Thus penalization can be extended to a general version for each $g_j$, the penalization of each gallery image $g_j$ is rewritten as:

$$w'_s(g_j) = \sum_{k=1}^{k_+} \log \{r(p, g_j^+) \cdot r(q_+, g_j)\} \quad (3)$$

Figure 2: An illustration of our conductive similarity and insulative dissimilarity. The preliminary experiment is conducted by KISSME [5]. 316 image pairs are randomly selected from the VIPeR dataset [3] for testing. Note that the distances between true matches and the strongly similar samples are significantly lower than those of strongly dissimilar samples.
where $r(p, g'_j)$ is set as $k_+$ if $g_j$ is not a member of $Q_+(p)$. It can be seen that the one which is more likely to be the true match is assigned with a smaller penalization and vice versa. Therefore, the ranking list is refined by:

$$r'(g_i) = r(g_i) - r(g_j)$$

(4)

where $r(g_i)$ denotes the rank of $g_i$ after the second step re-ranking. In this way, the refined ranking list via conductive similarity is achieved.

2.2 Dissimilarity Ranking Optimization

The similarity ranking optimization do improve the quasi-similar samples’ ranking orders. Then we penalize the quasi-dissimilar samples in dissimilarity ranking optimization. It also can be divided into three steps: First, the strongly dissimilar samples are achieved by the bottom-$k$ ($k_-$) results of a baseline method. Second, the strongly dissimilar samples are treated as new probes to re-query in the original gallery set. Third, a quasi-dissimilar set is achieved by backward re-query of the strongly dissimilar samples, and then penalize the ranking orders of the quasi-dissimilar samples due to their frequency in the quasi-dissimilar set. The details are discussed as follows.

The bottom-$k$ ($k_-$) farthest neighbours of the original method are defined as the strongly dissimilar samples of the probe, which is denoted as $G_-(p) = \{g'_j \mid j = 1, 2, \ldots, k_\}$ as the strongly dissimilar samples of image $p$. $k_-$ is the number of the strongly dissimilar samples. Similarly, the strongly dissimilar samples $G_-(p)$ are chosen for backward re-query. In other words, we treat each $g'_j$ in $G_-(p)$ as a new probe to search in the original gallery set. After the backward re-query, we’ll get $k_-$ ranking lists, and each top-$k$ of the backward ranking list can be expressed as $N_{k_+}(g'_j)$. Take the quasi-dissimilar union set of top-$k$ results $Q_-(p) = \{N_{k_+}(g'_j) \mid j = 1, 2, \ldots, k_\}$. Similarly, the general penalization of each $g_i$ is computed by:

$$w^+(g_i) = \sum_{k=1}^{k} \log\{r(p, g'_j - N + k_\} \cdot (k_- - r(g'_j, g_i)) + 1\}$$

Note that $r(g'_j, g_i)$ is set as $k_-$ if $g_i$ is not a member of $Q_-(p)$. $N_{k_+}(g'_j)$, and $r(p, g'_j) = N - k_-$ if $g_i$ is not an element of $Q_-(p)$. It can be seen that a higher penalization value is assigned for the one which is more likely to be wrong match. Therefore, the ranking list is refined by:

$$r'(g_i) = r(g_i) - w^+(g_i)$$

(6)

where $r(g_i)$ denotes the original rank of $g_i$. Specially, $r(g_i)$ can also be the revised ranking order after the similarity ranking optimization.

2.3 Complexity Analysis

Some implementation details are discussed in this subsection. As can be seen from above descriptions, the majority of the computation is spent on deriving the backward re-query ranking lists, which consist of mass pair-wise similarity computation between gallery images. In some special practical applications, e.g. surveillance video investigation, the gallery images can be obtained before querying the probe image. Therefore, the computation of our method is mainly spent off-line, where the similarity cues and dissimilarity cues can be computed offline together. In the off-line part, all gallery images are mutually compared with a computation complexity $O(n(n-1)/2)$ and a ranking complexity of $O(n^2 \log n)$. In the online portion, it only needs to compute the distance between the probe and every gallery image whose computation complexity is $O(n)$ and ranking complexity is $O(n \log n)$. With the above two-phrase implementation manner, the whole algorithm’s complexity can be greatly reduced and its online part is only proportional to the size of the top-$k$ and bottom-$k$, which is suitable in some practical applications.

3. EXPERIMENTAL RESULTS

3.1 Datasets and evaluation protocol

We evaluate our method on two publicly available datasets, the VIPeR dataset [3] and the CUHK01 dataset [14]. We chose these datasets that they provide many challenges faced in practical surveillance, i.e., viewpoint, pose and illumination changes, different backgrounds, low image resolutions, occlusions, etc. The VIPeR dataset contains 632 person image pairs captured from two different static camera views in outdoor academic environment. The dataset is challenging due to the intensive viewpoint and illumination changes. All the images are normalized to 128×48 for experiments. The CUHK01 dataset is also obtained from two disjoint camera views in an outdoor campus environment. It contains 971 persons with 3,884 images, and each person has two images in each camera. The person images in camera A are mostly captured by frontal view or back views while camera B captures the side views. All the images are normalized to 160×60 for experiments.

All quantitative results are exhibited in Cumulated Matching Characteristics (CMC) curves [11]. The CMC curve is a plot of the recognition performance versus the rank score and represents the expectation of finding the correct match inside top $k$ matches. Following the evaluation protocols described in [5,20], we randomly partition the dataset into two even parts, 50% for training and 50% for testing, without overlap on person identities. Especially, gave a comparison to two representative methods: multi-feature matching based on a classic metric learning method LMNN [15] and KISSME [5] as for its higher speed for testing.

3.2 Evaluation

Parameter Analysis Two important parameters $k_+$ and $k_-$ are analyzed in this section. Note that the baseline method is the original KISSME [5] conducted on VIPeR dataset. And CMC@01, CMC@05 and CMC@15 are reported for different $k_+$s and $k_-$s as shown in Fig 3. Note that the number of selected image pairs is 316. For $k_+$, the peak is somewhere around $k_+ = 60$. It is evident from Section 2.1 that smaller value of $k_+$ indicates a slighter remodification with limited improvement, but when $k_+$ is too large which may lead unreliable strongly similar samples, thus caused negative effects. Moreover, the curves drop earlier for CMC@01 rather than CMC@15, it can further illustrate the conductive similarity. For $k_-$, different $k_-$s have slight
influence on CMC@1 and CMC@5 due to the insulative dissimilarity mainly focus on improving the ranking orders of quasi-dissimilar samples. And the best choice for $k_-$ is 60.

Effectiveness To verify the effectiveness of our proposed method, three baseline methods are conducted on VIPeR dataset as shown in Fig.4 (a-c) and another experiment on CUHK01 dataset is shown in Fig.4 (d). And the parameters are set $k_+ = 60$ and $k_- = 60$ in all the experiments. Several conclusions can be drawn from the above figures: (1) Conductive similarity can truly improve matching results by pulling the ranking orders of quasi-similar samples. As shown in the figures, our “Similarity” has 5-10% improvements compared to the baseline method. (2) Insulative dissimilarity ameliorates the results by pushing the ranking orders of quasi-dissimilar samples. Specially, according to the growth rate of the CMC curves, the “Similarity” mainly focus on the improving former part of the rank which corresponds to the quasi-similar samples, while the “Dissimilarity” focus on the middle part with a higher increasing speed which corresponds to the quasi-dissimilar samples. (3) Combination of the two cues truly improves the matching results further.

Moreover, we compare our performance (based on KISSME + SCNCD) in the range of first 25 ranks to some existing state-of-the-art methods as shown in Table 1. It indicates that our final results compare favorably with state-of-the-art methods.

4. CONCLUSION

In this paper, we address ranking optimization approaches for person re-identification problem, a novel and efficient ranking optimization method based on similarity and dissimilarity is conducted. The main idea is that the correct match should be similar to the probe’s strongly similar samples and dissimilar to the strongly dissimilar samples, we pull the quasi-similar samples and push the quasi-dissimilar samples to optimize the ranking list. Extensive experiments without any parameters learning on two publicly data sets have validated the effectiveness. In the future, we will further estimate our method to combine multiple ranking lists and investigate intelligent ranking learning.

Table 1: Top ranked matching rates [%] on VIPeR.

<table>
<thead>
<tr>
<th>Rank →</th>
<th>( r = 1 )</th>
<th>( r = 5 )</th>
<th>( r = 15 )</th>
<th>( r = 25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELF [4]</td>
<td>12.08</td>
<td>31.28</td>
<td>54.00</td>
<td>65.00</td>
</tr>
<tr>
<td>SDALF [2]</td>
<td>19.87</td>
<td>38.89</td>
<td>58.22</td>
<td>70.00</td>
</tr>
<tr>
<td>PRDC [21]</td>
<td>15.66</td>
<td>38.42</td>
<td>64.00</td>
<td>72.78</td>
</tr>
<tr>
<td>KISSME [5]</td>
<td>22.63</td>
<td>50.13</td>
<td>71.65</td>
<td>82.12</td>
</tr>
<tr>
<td>SalMatch [19]</td>
<td>30.16</td>
<td>52.00</td>
<td>73.41</td>
<td>*</td>
</tr>
<tr>
<td>MidFeat [20]</td>
<td>29.11</td>
<td>60.08</td>
<td>85.78</td>
<td>91.28</td>
</tr>
<tr>
<td>SCNCD [16]</td>
<td>37.72</td>
<td>68.45</td>
<td>87.06</td>
<td>93.80</td>
</tr>
</tbody>
</table>

Our_Both | 40.98     | 71.54     | 89.68     | 94.08     |

Figure 4: Effectiveness of our method. “Similarity” and “Dissimilarity” denotes the similarity ranking optimization and dissimilarity ranking optimization respectively. “Both” indicates combine them together. (a) LMNN [15] on VIPeR; (b) Original KISSME [5] on VIPeR; (c) KISSME + SCNCD [16] on VIPeR; (d) KISSME on CUHK01 dataset.

Acknowledgement. The research was supported by National Nature Science Foundation of China (61303114, 61231015, 61170023), the Specialized Research Fund for the Doctoral Program of Higher Education (201304112024), the Technology Research Project of Ministry of Public Security (2014JSYJA016), the China Postdoctoral Science Foundation funded project (2013MS530350), the major Science and Technology Innovation Plan of Hubei Province (2013AA020), the Guangdong-Hong Kong Key Domain Breakthrough Project of China (2012A090200007), and the Special Project on the Integration of Industry, Education and Research of Guangdong Province (2011B090400601), Nature Science Foundation of Hubei Province (2014CBF72). Jiangxi Youth Science Foundation of China(20151BAB217013).

5. REFERENCES